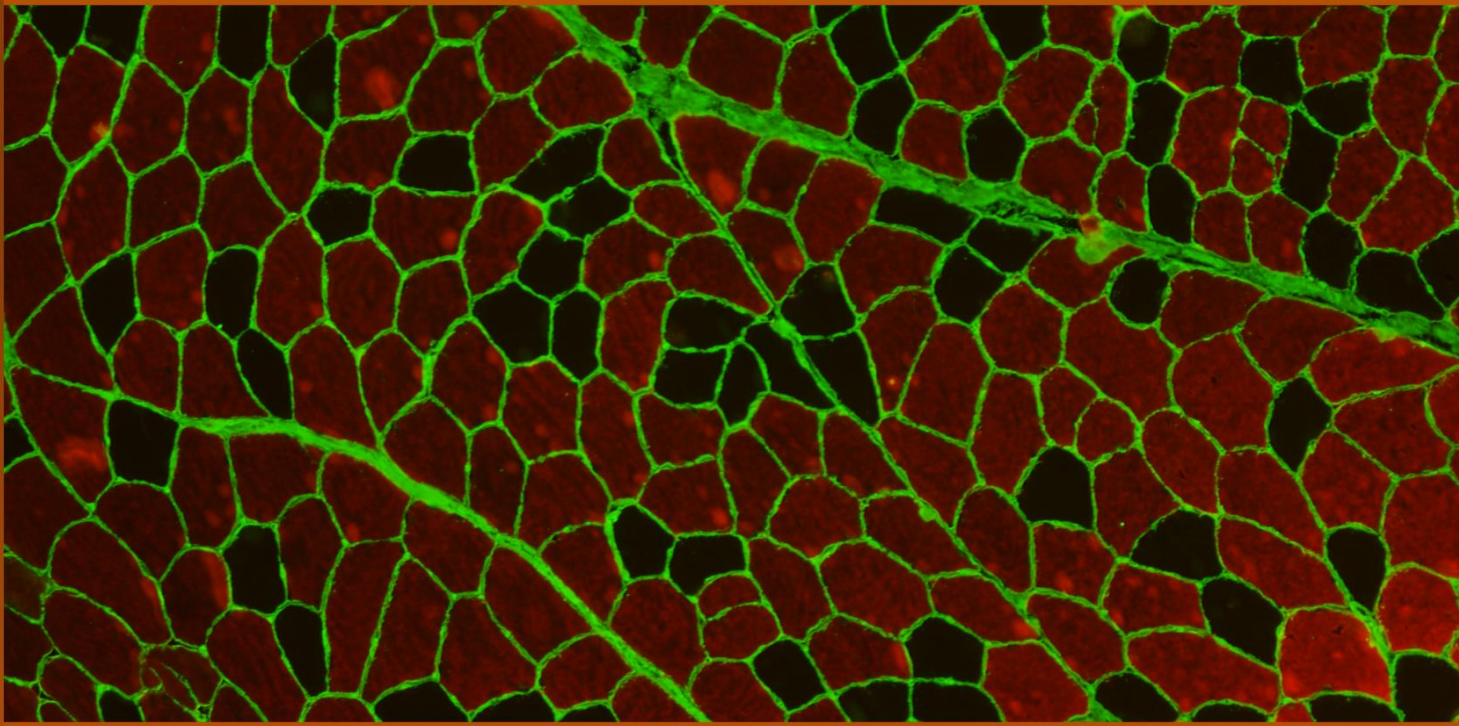
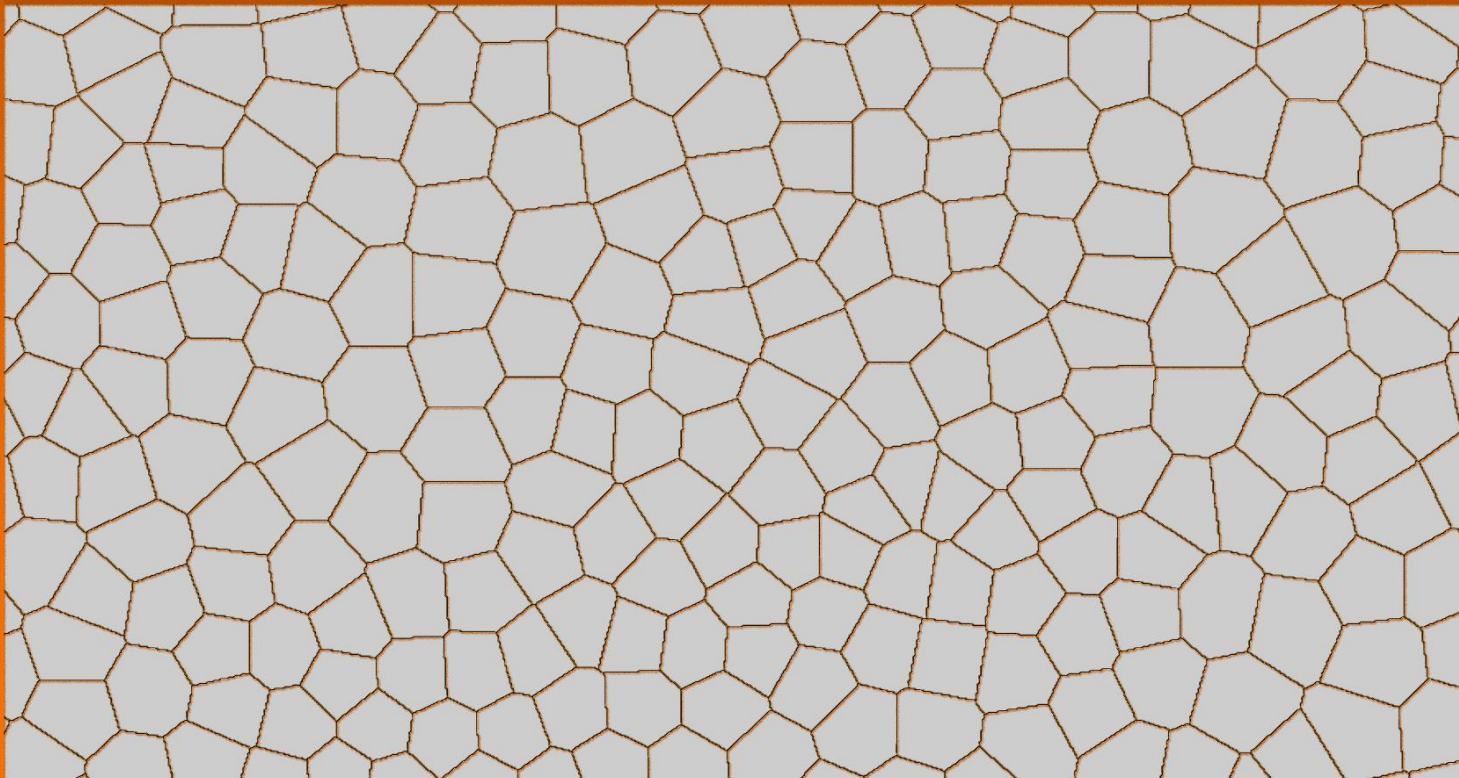


• Daniel Sánchez Gutiérrez •



Estudio de la organización compleja de tejidos biológicos usando conceptos matemáticos

Tesis Doctoral



Universidad de Sevilla

Facultad de Biología

Dpto. de Biología Celular



**Estudio de la organización compleja
de tejidos biológicos usando
conceptos matemáticos**

Trabajo presentado para optar al grado de Doctor
por la Universidad de Sevilla

Daniel Sánchez Gutiérrez
SEVILLA, 2015

Doctorando



Daniel Sánchez Gutiérrez

Director



Luis M. Escudero Cuadrado

Tutor



Manuel Muñiz Guinea

“Todos los elementos de tu cuerpo fueron forjados hace muchos, muchos millones de años en el corazón de una estrella lejana que explotó y murió. Esa explosión esparció esos elementos a través de la desolación de ese espacio profundo. Después de muchos, muchísimos millones de años, esos elementos se unieron para formar nuevas estrellas, nuevos planetas, y así siguieron sin cesar. Los elementos se unieron y se separaron formando zapatos, y barcos, y lacre, y repollos, y reyes, hasta que al final se unieron para crearte a ti. Eres único en el universo”

- El Doctor -

Agradecimientos

Cuando realicé la carrera de Ingeniería en Telecomunicaciones jamás pensé en realizar el Doctorado y mucho menos en Biología. Todo comenzó cuando mis tutoras del proyecto fin de carrera Begoña y Carmen me pusieron en contacto con un tal Luisma, un biólogo que buscaba generar “paper como churros” empleando conceptos tan novedosos como frikis. Este fue el punto de partida de un trabajo que ha supuesto un gran reto y que no hubiera sido posible sin la intervención en mayor o menor medida de un gran número de personas. Todos ellos, merecen mis más sincera admiración y respeto.

En el laboratorio 109 encontré a un grupo de personas incansables con un humor insuperable, una energía extrema y que compartían una pasión en común por la ciencia. Gracias a todos vosotros por hacerme participe de vuestros avances y progresos, por hacerme conocedor de vuestros pensamientos e inquietudes.

Gracias María por ver el lado práctico de las cosas y por haber sido una fantástica compañera durante los pocos meses que compartimos espacio. Tu marcha fue una gran pérdida, y espero que durante el tiempo que he ocupado tu sitio se me haya pegado algo de la fuerza con la que afrontas el día a día.

Gracias Rosana, por dar el punto de cordura al laboratorio. Siempre organizada, siempre con las ideas claras. Gracias por parar tu trabajo cada vez que necesitaba comprender algún que otro concepto básico y siempre haciéndolo con una sonrisa en la cara.

Gracias Antonio (alias tresmi) por las innumerables horas de entretenimiento que has brindado con tu música, discursos políticos y con tus imitaciones. Has sido un gran compañero.

Gracias Clara por tu entereza, tu saber estar, tu buen humor y tu inigualable actitud ante el trabajo. Has sido mi punto de referencia y has hecho que las inagotables horas en el laboratorio sean mucho más amenas. En mi memoria se quedan frases como “*Ponerte el chándal del Decathlon no lo convierte en deporte*”, “*¿Te ha pintado un cubista? Ah no, estas enfadá*” o “*Nuca me refiero a mí misma,....Yo, como Clara que soy*”. Has sido una trabajadora incansable, haciendo bien todo lo que te proponías (obviando el detallito del Western Clara), me siento muy agradecido de haber podido trabajar a tu lado.

Gracias Dori por ser el alma del laboratorio. Tú llevas el control y eres sabia conocedora de todo lo que pasa dentro de esas cuatro paredes. Te preocupas porque todo el mundo esté bien y siempre estás dispuesta a echar una mano a quien lo necesite. Has sido mi confidente y una grandísima compañera. Aunque no puedo compensarte con un “mini cerdito enano volador” por todo lo bueno que me has dado, puedo darte las gracias una y mil veces. ¡Gracias!

Manolo para ti las palabras de agradecimiento se me quedan cortas. Llegaste al laboratorio en un momento tenso y complicado para mí. Tu compañía y tu espíritu de trabajo fue inspirador y me ayudó a renovar energías. Tu compañía en las extensas horas de laboratorio compartiendo risas no tienen precio. Desbordas paciencia, responsabilidad y una atención al detalle digna de estudio (de otra forma no se comprende el enorme trabajo que realizas con las moscas). Por todo ello, ¡Muchas Gracias!

Alberto, el jefe, el compañero. Responsable máximo de todo lo que ocurre en el laboratorio. Generoso, exigente y un grandísimo profesional y aunque en ocasiones seas algo despistadillo, no tengo dudas que sin ti, nada de lo que acontece en el 109 podría ser posible. ¡Muchísimas gracias por la labor que desempeñas y la energía con la que lo haces!....solo decirte una cosa... ¡VORONOI!!!!

Gracias Luisma por ser un jefe y director de tesis tan poco convencional. De los que hacen el avioncito por los pasillos cuando las cosas marchan, de los que están siempre a tu disposición para cualquier problema o idea que surja y de los que te ponen videos para relajar el ambiente. Este trabajo no podría haberse dado sin ti, por tu cabezonería, por tu perfeccionismo, por tu paciencia y tu impaciencia, por tu dedicación, por tu persecución hasta fregando los platos en la cocina para saber cómo va la simulación por todo esto y por mucho más ¡MUCHAS GRACIAS!.

En el Departamento de Biología Celular he pasado el último año de tesis, donde Manolo, José Manuel, Inma, Paula, Ángel, Luis, Valle, Marga y Marcos me han acogido y han hecho mi estancia muy agradable. Gracias a todos.

Gracias Reme por ser siempre tan atenta y velar por nosotros. Muchas gracias por tus comidas, por compartir tus historias, por dedicarnos tiempo, por escucharnos, por animarnos pero sobretodo por trasmitirnos tu alegría. Gracias a ti nunca volveré a ver una piedra con los mismos ojos.

Ana Mari, Laura y Pedro... ¡que puedo decir de vosotros! ¡Sois los mejores! Que grupo más variopinto nos hemos juntado en el departamento. Nos veíamos en visitas furtivas pero nos explayábamos sobretodo en la hora de comer y qué comidas! Ana Mari tu eres el símbolo de la alegría y el dinamismo (lo mismo realizas chorrocintas tareas de laboratorio tú sola que bailas reggaetón), Laura tu perseverancia y nivel de perfeccionismo es envidiable y Pedro... ¡tú eres simplemente un máquina! Siempre he encontrado una mano tendida en vosotros ¡Muchísimas Gracias!

Gracias a todos mis amigos, pero especialmente a Ana, Elena, Goretti, Oscar, Juanma, Javi, Víctor, Jorge, Rafa, Silvia, Fran, Meu y Mari. Vosotros sabéis como ponerle buena cara a la vida.

Gracias, muchísimas gracias a Elena, mi Elena. Esta tesis ha ido creciendo junto contigo, al principio de forma dudosa y cogiendo fuerza y consistencia con el paso del tiempo. Gracias por escucharme y entenderme, gracias por apoyarme y comprenderme. Esta tesis lleva mucho de ti.

Gracias a mi familia, especialmente a mis padres, Ángel y Carmen, y a mi hermano Guillermo. Durante el desarrollo de este trabajo han cambiado un montón de cosas, quizás demasiadas, pero sé que nunca va a cambiar vuestro apoyo incondicional. Papa, Mama muchas gracias por apoyarme en todo, vosotros habéis sido los principales sufridores. Guille, hermano, a ti te ha tocado la peor parte, has tenido que sufrir mi despertador, mis cambios de ánimo, mi ausencia... siendo eso solo la punta del iceberg. A vosotros os agradezco especialmente y os dedico el trabajo realizado en esta tesis.

GRACIAS

Índice general

Índice de figuras.....	III
Índice de tablas.....	IV
Abreviaturas	V
Prólogo.....	VII
Introducción.....	1
1.Patrones geométricos en la naturaleza.....	3
2.Tejidos epiteliales.....	4
2.1.Estructura del tejido epitelial.....	4
2.2.El disco imaginal de ala de <i>Drosophila</i> como modelo epitelial.....	5
2.3.Desarrollo y crecimiento epitelial	7
3.Tejidos musculares.....	8
3.1.Estructura del musculo esquelético	9
3.2.Desarrollo y crecimiento del músculo esquelético	11
4.La geometría como una herramienta para entender la biología de los tejidos.....	11
4.1.Propiedades geométricas en tejidos empaquetados.....	12
4.2.Patrones geométricos	12
Objetivos.....	15
Materiales y Métodos	19
1.Materiales	21
1.1.Tejidos empaquetados proliferativos	21
1. 2.Tejidos empaquetados no proliferativos	22
2.Extracción de características de imágenes naturales	22
3.Evaluación de diferencias entre 2 categorías de imágenes	27
3.1.Comparación de células individuales: Correlación	27
3.2.Comparación de histogramas de frecuencia de polígonos: MANOVA.....	29
3.3.Comparación de la organización tisular entre dWL y dWP	29
3.4.Aleatorización de imágenes	31
4.Simulación de tejidos naturales empaquetados.....	32
4.1.Diagramas de Voronoi y vía CVT	32
4.2.Simulación de tejidos biológicos por ordenador.....	34
5.Análisis estadístico de la distribución del área celular en tejidos naturales y diagramas de Voronoi.....	36
6.Selección de subconjuntos de células.....	37

6.1.Células con atrofia neurógena	37
Resultados	39
1.Estudio de la influencia de la proliferación en la organización de tejidos epiteliales.....	41
1.1.Análisis de características celulares en muestras de larva y prepupa en el disco imaginal de ala de <i>Drosophila</i>	41
1.2.Estudio de la distribución de polígonos en el epitelio del ala en <i>Drosophila</i> durante el desarrollo larvario	44
1.3.Estudio de características de grafos como medida de la organización del epitelio durante el desarrollo larvario.....	45
1.4.Relación entre los cambios en la distribución de polígonos y la aparición de una organización epitelial diferente.	48
2.Restricciones físicas de las células como mecanismo regulador de la organización de los tejidos empaquetados.....	51
2.1.Análisis de la distribución de polígonos en diversos tejidos empaquetados.....	51
2.2.Análisis de tejidos mediante el uso de conceptos geométricos	51
2.3.Restricciones geométricas en los tejidos empaquetados.....	56
2.4.Restricciones biofísicas en los tejidos empaquetados	61
2.5.Propiedades biofísicas que dirigen la organización de los tejidos naturales empaquetados	64
Discusión	69
1.Influencia de la proliferación en la organización de tejidos epiteliales	71
2.Restricciones físicas de las células como mecanismo regulatorio de la organización	74
Conclusiones	79
Bibliografía	83
Anexo	93
Anexo I: Funciones polinómicas de ajuste a la vía CVT	94
Anexo II: Software suplementario.....	96
Anexo III: Artículo publicado correspondiente al capítulo I de resultados	101

Índice de figuras

Fig. 1. Ejemplos de teselaciones presentes en la naturaleza.	3
Fig. 2. Estructura del tejido epitelial.....	5
Fig. 3. Organización disco imaginal de ala de <i>Drosophila</i>	6
Fig. 4. Representación del ciclo de vida de <i>Drosophila</i>	8
Fig. 5. Organización de un musculo esquelético	10
Fig. 6. Imágenes epiteliales y correlación entre células..	42
Fig. 7. Distribución de polígonos en dWL y dWP.....	45
Fig. 8. PCA de imágenes dWL y dWP	46
Fig. 9. Descriptor de PCA para 4000 combinaciones aleatorias de imágenes en 2 grupos	48
Fig. 10. Randomización de combinaciones de imágenes y <i>p</i> value del test MANOVA para sus respectivas distribuciones de polígonos.....	49
Fig. 11. Representación del descriptor PCA frente al <i>p</i> value del test MANOVA para distintas combinaciones de imágenes	50
Fig. 12. Distribución de polígonos de tejidos biológicos comparados con diagramas de Voronoi	52
Fig. 13. Comparación de la distribución de polígonos de tejidos naturales con la vía CVT.....	55
Fig. 14. Leyes geométricas que definen a los tejidos empaquetados y diagramas de Voronoi..	57
Fig. 15. Distribución de área de tejidos naturales y geométricos	59
Fig. 16. Comparación distribución de área de tejidos alterados con tejidos naturales y diagrama de Voronoi.....	62
Fig. 17. Distribución de polígonos de tejidos alterados.	63
Fig. 18. Alteraciones biofísicas en tejidos proliferativos usando simulación por ordenador.	65
Fig. 19. Distribución de polígonos de diferentes subconjuntos de células provenientes de muestras de músculos y simulaciones.	68
Fig. 20. Simulación por ordenador muestra que homogeneidad en el volumen de estabilidad regula las constricciones del tejido.....	67
Fig. 21. Modelo regulatorio de la organización del tejido epitelial durante el final de la fase larvaria proliferativa.	73
Fig. 22. Esquema conceptual de la organización de tejidos empaquetados.....	76

Índice de tablas

Tabla 1. Lista de características analizadas	43
Tabla 2. Imágenes y células analizadas en las categorías dWL y en dWP	44
Tabla 3. Valores de las características seleccionadas	47
Tabla 4. Distribución de polígonos de tejidos naturales y de patrones geométricos	53
Tabla 5. Comparación tejidos naturales con CVT continuo	56
Tabla 6. Teorema de Euler	58
Tabla 7. Estadísticos y percentiles de la distribución de área	60

Abreviaturas

Afect.	Afectadas
Atrof.	Atróficas
BC	Células del borde (B order C ells)
BCA	Bíceps control adulto (B iceps C ontrol A dult)
BNA	Bíceps con atrofia neurógena adulto (B iceps N eurogenic A trophy)
cc.	Características
cDP	Células columnares del disco propio (c olumnar D isc P roper)
cNT	Neuroepitelio del pollo (c hicken N eural T ube epithelium)
cols.	Colaboradores
Contr.	Control
CVT	Diagrama de Voronoi Centroidal (C entroidal V oronoi T essellation)
D	Diagrama
D.E.	Desviación estándar
dWL	Ala de <i>Drosophila</i> en estadio de larva (D rosophila W ing L arva)
dWP	Ala de <i>Drosophila</i> en estadio de prepupa (D rosophila W ing P repupa)
dMWP	Ala mutante de <i>Drosophila</i> en estadio de prepupa (D rosophila M utant W ing P repupa)
DP	Disco propio (D isc P roper)
EYE	Disco de ojo de <i>Drosophila</i>
Fig.	Figura
MANOVA	Análisis Multivariado de la Varianza (M ultivariate A nalysis O f V ariance)
mDP	Células marginales del disco propio (m arginal D isc P roper)
mPE	Células marginales del epitelio peripodial (m arginal P eripodial E pithelium)
NBC	Células fuera del borde (N ot B order C ells)
PCA	Análisis de componentes principales (P rincipal C omponent A nalysis)
PE	Epitelio peripodial (P eripodial E pithelium)
Per.	Percentil
Prol.	Proliferación
ROI	Región de interés (R egion O f I nterest)
SEM	Error estándar de la media (S tandard E rror of the M ean)
Sim.	Simulación
sPE	Células escamosas del epitelio peripodial (s quamous P eripodial E pithelium)
VC	Células válidas (V alid C ells)

Prólogo

Esta tesis doctoral se basa en un importante trabajo previo de procesamiento de imágenes llevado a cabo en el laboratorio del Dr. Luis M. Escudero. Así mismo quiero destacar la cesión del algoritmo de tratamiento de imágenes para la segmentación de tejidos musculares desarrollado por la Dra. Aurora Sáez del grupo de la Dra. Carmen Serrano y la Dra. Begoña Acha de la Escuela Técnica Superior de Ingeniería de la Universidad de Sevilla. Estas muestras musculares fueron facilitadas por Eloy Rivas y la Dra. Carmen Paradas del Hospital Virgen del Rocío de Sevilla.

Quiero hacer constar mi agradecimiento a varios investigadores, entre los que hay que destacar a la Dra. Melda Tozluoglu del laboratorio de la Dra. Yanlan Mao (University College London) por los experimentos en los que se simulan tejidos empaquetados (capítulo II), al Dr. Joseph Barry (European Molecular Biology Laboratory) y al Dr. Alberto Pascual (Instituto de Biomedicina de Sevilla). Los experimentos que se presentan, así como el planteamiento de la línea de investigación han sido posibles gracias a la colaboración de todos ellos.

Introducción

1. Patrones geométricos en la naturaleza

En la naturaleza podemos encontrar infinidad de patrones geométricos naturales tales como fractales, espirales o teselaciones, que han intrigado a matemáticos y biólogos durante décadas (Haeckel, 1904; Stevens, 1974). Las teselaciones representan un tipo de patrón geométrico compuesto por figuras poligonales que rellenan un plano sin solaparse de tal forma que no queden espacios entre ellos. Estos patrones pueden hallarse en diversos tejidos celulares tanto animales como vegetales, en las marcas que dibuja la piel de una jirafa o incluso en formaciones geológicas volcánicas (Fig. 1).

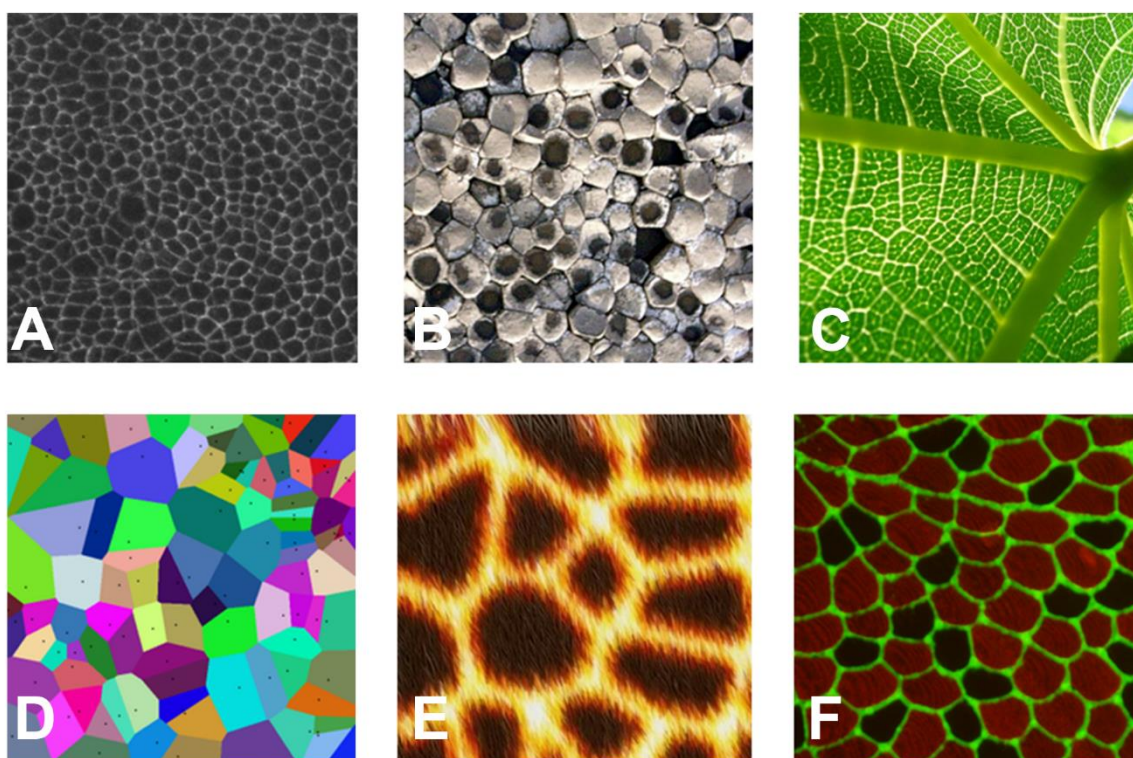


Fig. 1. Ejemplos de teselaciones presentes en la naturaleza. A) Teselación en el tejido epitelial del disco de ala de *Drosophila*. B) Teselación en la composición de rocas basálticas (Calzada del gigante). C) Teselación en el tejido vegetal. D) Teselación geométrica en un diagrama de Voronoi. E) Teselación en la piel de una jirafa. F) Teselación en el tejido muscular esquelético de bíceps humano.

En los seres vivos las células se organizan en tejidos. Las células de un mismo tejido se comunican entre sí por medio de uniones intercelulares especializadas, lo que facilita la colaboración entre ellas y permiten que operen como una unidad funcional. Una asociación precisa de los tejidos da como resultado el funcionamiento adecuado de cada órgano. En esta tesis profundizaremos en los tejidos biológicos empaquetados. El tipo de empaquetamiento celular en estos tejidos va a determinar su patrón morfológico. Estudiar tejidos biológicos empaquetados empleando conceptos

geométricos, tratando a las células como polígonos, ha sido ampliamente utilizado para investigar la arquitectura del tejido y como se remodela durante el desarrollo (Chichilnisky, 1986; Classen y cols., 2005; Hayashi y Carthew, 2004; Honda, 1978; Rivier y Schliecker, 1995).

En este trabajo usaremos como modelos los tejidos epiteliales compuestos por células poliédricas que revisten superficies y tejidos musculares compuesto por células alargadas o fibras con capacidad de contraerse (Ross y Wojciech, 2012; Junqueira y Carneiro, 2015).

2. Tejidos epiteliales

El tejido epitelial es una de las unidades fundamentales del desarrollo de un animal. Esta clase de tejido presenta como función principal el revestimiento de las cavidades y las superficies internas y externas de los órganos o del cuerpo en su totalidad. Normalmente, esta función se asocia con otras tales como protección, absorción de iones y moléculas (riñones e intestinos) y percepción de estímulos (neuroepitelios olfatorios y gustativos).

2.1. Estructura del tejido epitelial

A partir del epitelio formado durante la embriogénesis aparecen diferentes tipos celulares que se organizarán para dar distintas partes de un organismo adulto (Lecuit y Le Goff, 2007; Axelrod, 2006). Las células que integran los epitelios poseen tres características principales. En primer lugar, están dispuestas unas muy cerca de otras debido a la falta de presencia significativa de sustancia intercelular o fibras. Las células se adhieren entre sí por medio de moléculas de adhesión celular específicas. A estas células se le asocian elementos del citoesqueleto como la actina o la miosina II. La miosina II también está involucrada en el proceso de citocinesis en la división celular (Conti y cols., 2008; Franke y cols., 2010). En segundo lugar, las células presentan polaridad morfológica y funcional. Las diferentes funciones celulares se asocian con tres regiones distintas (región apical, región lateral y región basal). En último lugar, presentan una superficie basal adherida a una membrana basal subyacente que es una capa de matriz extracelular de sostén (Fig. 2A), (Ross y Wojciech, 2012).

Los tejidos epiteliales se pueden clasificar según la cantidad de estratos celulares, en simple (si el tejido posee una sola capa celular) o estratificado (si el tejido presenta más de una capa). Dentro de los epitelios monocapa, se considera que un epitelio es

pseudoestratificado cuando los núcleos están dispuestos a diferentes niveles. También se puede clasificar atendiendo a la forma de las células que lo componen, en planas (escamosas), cúbicas (cuboidales) o cilíndricas (columnares), (Fig. 2B,C), (Ross y Wojciech, 2012).

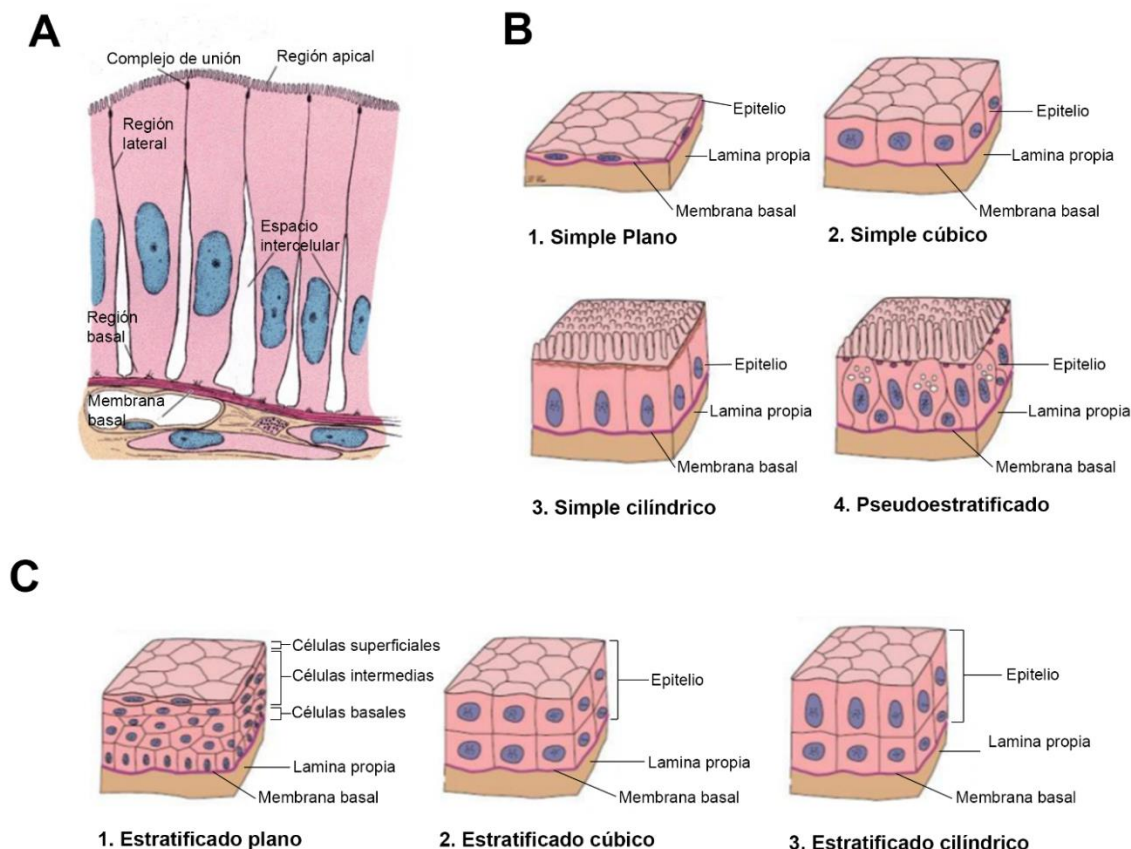


Fig. 2 Estructura del tejido epitelial. **A)** Diagrama de célula epitelial. En el diagrama se indican tres regiones de una célula epitelial típica. El complejo de unión provee adhesión entre las células contiguas y separa el espacio luminal del espacio intercelular. **B)** Tipos de epitelios de revestimiento simples y epitelio pseudoestratificado. Los distintos tipos de epitelio se apoyan sobre un tejido conjuntivo subyacente, ambos separados por una membrana basal. **C)** Tipos de epitelios de revestimiento estratificado. Los distintos tipos de epitelio se apoyan sobre un tejido conjuntivo subyacente, ambos separados por una membrana basal. Imágenes obtenidas y modificadas de: Ross y Wojciech, 2007 (**A**) y Cui, 2010 (**B** y **C**).

2.2. El disco imaginal de ala de *Drosophila* como modelo epitelial

Uno de los organismos modelos más importantes para el estudio de tejidos epiteliales es la mosca de la fruta, *Drosophila melanogaster*. Entre las primeras publicaciones con *Drosophila* se encuentra la realizada en 1910 por el Dr. Thomas H. Morgan (Morgan, 1910), desde entonces se han realizado innumerables investigaciones utilizando este organismo modelo, lo que ha mejorado nuestra comprensión de la biología, especialmente en el campo de la Genética y en la Biología

del Desarrollo. En la caracterización de un organismo como modelo está implícito el supuesto de que, al conocer su biología, ésta pueda explicar al menos en parte, la de otros seres vivos. Este organismo tan ampliamente estudiado ha sido validado como organismo modelo ya que, a pesar de ser un insecto, muchos de sus procesos celulares son similares a los de otros organismos superiores incluidos los humanos. Adicionalmente, comparando su mapa genómico con el de los humanos se han identificado que aproximadamente el 61% de los genes involucrados en enfermedades humanas que se conocen tienen una contrapartida identificable en el genoma de las moscas de la fruta (Adams y cols., 2000).

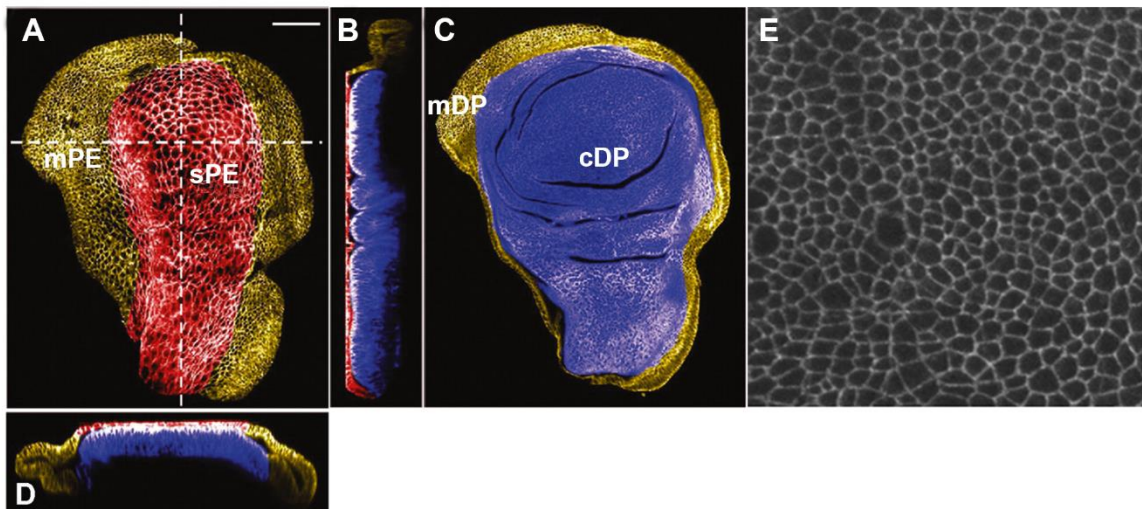


Fig. 3. Organización disco imaginal de ala de *Drosophila*. La figura muestra distintas secciones de un disco de ala pseudo-coloreado orientado con la parte anterior a la izquierda y dorsal hacia abajo. En rojo se muestran las células escamosas del epitelio peripoidal (sPE). En amarillo se muestran las células cuboidales del margen del epitelio peripoidal (mPE) y del disco propio (mDP). Líneas blancas discontinuas indican la orientación de la sección transversal y longitudinal. En azul se muestran las células columnares del disco propio (cDP). **A)** Imagen confocal mostrando una sección del PE. **B)** Sección longitudinal del disco en su parte central. **C)** Imagen confocal mostrando una sección del DP (5 μ m por debajo del PE). **D)** Sección transversal. **E)** Región de interés de un grupo de células columnares del DP. El contorno celular se visualiza por medio de la proteína Armadillo. Imágenes **A-D** obtenidas de McClure y Schubiger, 2005.

Las ventajas que *Drosophila* tiene para poder realizar investigaciones en biología celular son varias. Entre otras, un tamaño pequeño, un ciclo de vida corto (hechos que son relevantes para el análisis genético, ya que permiten cultivar un gran número de organismos en espacios reducidos y obtener nuevas generaciones en poco tiempo), un bajo coste, una enorme cantidad de herramientas genéticas producidas por la activa y prolífica comunidad investigadora y la posibilidad de realizar experimentos *in vivo* que permiten estudiar eventos dinámicos (Rodríguez-Arnaiz y cols., 1992; Jiménez y Merchant, 2003).

Gran parte de la investigación de *Drosophila* se realiza en el estadio de larva, en concreto en los discos imaginales, que son estructuras epiteliales larvarias que darán

la mayor parte de la mosca adulta. Los discos imaginales son sacos de células epiteliales que tienen su origen en grupos de células ectodérmicas indiferenciadas del embrión (Bate y Martínez-Arias, 1991). Durante los estadios larvarios, estas células, a diferencia del resto de células de la larva que no se dividen, proliferan exponencialmente (García-Bellido y Merriam, 1971). Como insecto holometábolo, la larva de *Drosophila* presenta en su ala un disco imaginal formado por 2 epitelios contiguos de diferente morfología, el epitelio peripodial monoestratificado (PE, del inglés “Peripodial Epithelium”) principalmente compuesto por células escamosas y el disco propio pseudoestratificado (DP, del inglés “Disc Proper”) compuesto principalmente por células columnares. Existe una tercera clase, células cúbicas, situadas en el borde del disco de ala que forman una zona de transición entre la capa columnar y el epitelio escamoso (Fig. 3A-D), (Aldaz y cols, 2010; Aldaz y cols., 2013). Numerosas investigaciones se han centrado en el DP por varias razones. En primer lugar, el epitelio columnar es mucho más visible que el PE y lo que es más importante, durante la morfogénesis, las células cilíndricas del disco representan la amplia mayoría del epitelio adulto mientras que las células peripodiales hacen tan solo una insignificante contribución. Como resultado de esto, el disco imaginal de *Drosophila* es a menudo descrito como epitelio columnar pseudoestratificado aunque este término no refleja exactamente su anatomía (Fig. 3E), (Aldaz y cols., 2010; Gibson y Schubiger, 2001).

2.3. Desarrollo y crecimiento epitelial

En términos morfogenéticos, el disco imaginal del primordio de ala ha sido muy bien estudiado para conocer qué factores regulan su desarrollo. Esto lo hace perfecto para analizar el desarrollo de la topología epitelial. En la Fig. 4 se presenta el ciclo de vida de *Drosophila*. Durante 4 días de crecimiento de la larva, un disco imaginal de ala pasa de estar formado por 20 células a tener aproximadamente 50000 (Milan y cols., 1996; Martín, Herrera y Morata, 2009). Posteriormente en el estadio de pupa, la división celular casi desaparece, llegando a formar un epitelio formado casi totalmente por células hexagonales que conformará el ala adulta tras la metamorfosis. Es de interés señalar que a través de este periodo de intensa proliferación la topología del epitelio parece heterogénea sin un orden o regla aparente que domine su organización. Estudios muy diversos han intentado comprender los mecanismos de organización de las células durante el desarrollo del tejido epitelial. Todos ellos, se basan directa o indirectamente en el número de lados de las células (Lecuit y Le Goff, 2007; Axelrod, 2006; Farhadifar y cols., 2007; Gibson y cols., 2006; Gho, Bellaïche y

cols., 1999; Classen y cols., 2005; Martínez-Arias y Stewart, 2002; Patel y cols., 2009; Aegerter-Wilmsen y cols., 2010; Gibson y Gibson, 2009; Eaton y Jülicher, 2011). Actualmente, esta aceptada la idea de una distribución poligonal que se conserva en los epitelios proliferativos de todos los metazoos (alrededor 3% de células de 4 lados, 28% de pentágonos, 46% de hexágonos, 20% de heptágonos y 3% de octógonos). Esta distribución estereotipada del número de lados de las células es una consecuencia matemática de la división celular a través de un análisis probabilístico utilizando la cadena de Markov (Axelrod, 2006; Gibson y cols., 2006; Gibson y cols., 2011), y en conjunción con los reordenamientos celulares del tejido (Aegerter-Wilmsen y cols., 2010). Esta distribución está empíricamente demostrada en muestras de tejidos de otros animales, como la epidermis de la rana *Xenopus* o la epidermis del cnidario de agua dulce *Hydra*. Estos estudios sugieren que la topología epitelial es irregular, pero no es azarosa debido a la existencia de una organización similar en todos los metazoos.

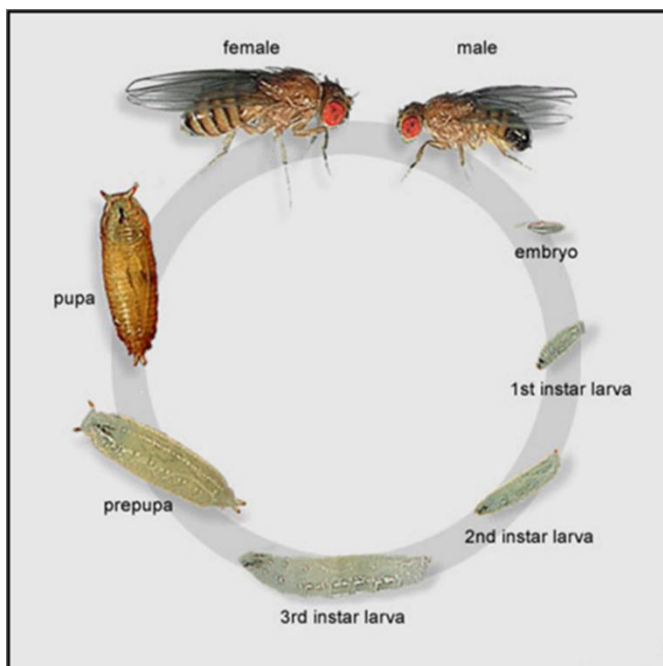


Fig. 4. Representación del ciclo de vida de *Drosophila*. Las etapas de desarrollo son embrión, larva (subdividida en 3 fases), prepupa y pupa. Imagen obtenida de la página FlyBase (www.flybase.org).

3. Tejidos musculares

El tejido muscular está formado por células alargadas que pueden contraerse y relajarse cuando son estimuladas permitiendo la realización de los movimientos del cuerpo, tanto del esqueleto como de sus órganos internos. Estas células están especializadas en la transformación de la energía en fuerza y/o movimiento. Esto sucede a través de la interacción molecular entre las proteínas motoras miosina y actina. Aun cuando en toda la escala zoológica el tejido muscular cumple básicamente con las características mencionadas, existe gran diversidad de células musculares,

sobre todo en los invertebrados. Nos centraremos en el estudio de tejidos musculares de vertebrados. Específicamente en los tejidos clasificados como muscular estriado esquelético. El músculo estriado esquelético está caracterizado por presentar células que exhiben estriaciones transversales. Además el tejido se fija a los huesos y tiene a su cargo el movimiento de los esqueletos axial y apendicular y el mantenimiento de la posición corporal (Gomez-Alvarez, 2007).

Uno de los músculos más utilizados para el estudio de tejidos musculares esqueléticos en anatomía humana es el bíceps braquial. El bíceps es un músculo largo y fusiforme de dos cabezas que se encuentra en la parte superior del brazo entre el hombro y el codo. Ambas cabezas surgen en la escápula y se unen para formar un tendón grueso muscular que se adjunta a la parte superior del antebrazo. Su función consiste en flexionar el codo y ayudar a la flexión de hombro.

3.1. Estructura del musculo esquelético

El músculo esquelético está compuesto por células con una forma generalmente alargada en el eje de contracción con una sección trasversal pequeña, por lo que habitualmente a estas células se les denomina fibras musculares. Estas fibras musculares estriadas se mantienen juntas mediante una fina capa de tejido conjuntivo. El tejido conjuntivo que rodea tanto a las fibras musculares individuales como los haces de fibras es indispensable para la transducción de fuerza. En los extremos de los músculos el tejido conjuntivo continúa en forma de tendón o de alguna otra estructura de fibras colágenas que sirve para fijarlos, la mayor parte de las veces, a los huesos (Gomez-Alvarez 2007; Ross y Wojciech, 2012; Junqueira y Carneiro, 2015). El tejido conjuntivo del músculo se designa según su relación con las fibras musculares. El endomisio es una banda muy delgada de tejido conjuntivo que rodea a las fibras musculares. El perimisio es una capa más gruesa del tejido conjuntivo que rodea a un grupo de células para formar un haz o fascículo. Los fascículos son unidades funcionales de fibras musculares que actúan en conjunto para desempeñar una función específica. El epimisio es la vaina del tejido conjuntivo denso que rodea todo el conjunto de fascículos que forman el músculo. La fibra muscular está repleta de subunidades de disposición longitudinal llamada miofibrillas, que a su vez, están compuestas por miofilamentos de 2 tipos: filamentos gruesos (de miosina) y filamentos finos (de actina). Las miofibrillas se extienden a lo largo de la célula muscular siendo su unidad funcional el sarcómero (Fig. 5), (Gomez-Alvarez 2007; Ross y Wojciech, 2007; Junqueira y Carneiro, 2015).

Esta clase de tejidos presentan células multinucleadas. Los múltiples núcleos están ubicados normalmente en la periferia celular. Existen otros núcleos, situados también en la periferia celular, inmediatamente por debajo de la lámina basal, que no forman parte de la célula muscular sino de células anexas denominadas satélites (Gomez-Alvarez 2007; Ross y Wojciech, 2012; Junqueira y Carneiro, 2015).

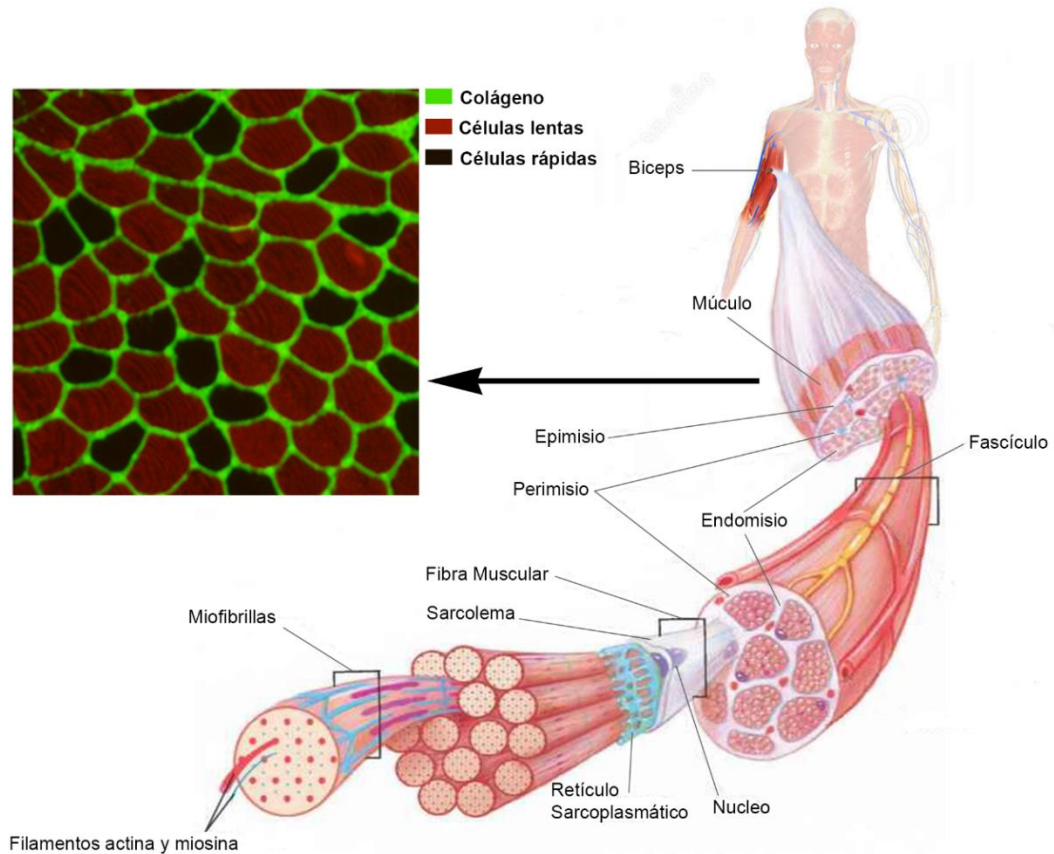


Fig. 5. Organización de un músculo esquelético. Un músculo esquelético está compuesto por haces de fibras llamados fascículos. A su vez, cada fascículo está formado por un conjunto de fibras musculares alargadas (células multinucleadas). La fibra muscular consiste en agrupaciones de unidades longitudinales, las miofibrillas, que a su vez están compuestas por miofilamentos de 2 tipo: filamentos gruesos (de miosina) y filamentos finos (de actina). Imagen modificada del libro Shier y cols., 1999.

Las fibras musculares esqueléticas se clasifican en tres tipos y lo habitual es que en cualquier músculo dado estén presentes los tres tipos de fibras; la proporción de cada tipo varía según la actividad funcional del músculo. Las fibras de tipo I o fibras oxidativas lentas son fibras pequeñas que forman unidades motoras de contracción lenta resistente a la fatiga pero generan menos tensión muscular que otras fibras. Las fibras de tipo IIa o fibras glucolíticas oxidativas rápidas son fibras de tamaño mediano y constituyen las unidades motoras de contracción rápida resistente a la fatiga que generan un gran pico de tensión muscular. En último lugar tenemos las fibras de tipo IIb o fibras glucolíticas rápidas que son fibras grandes que integran las unidades motoras de contracción rápida propensas a la fatiga y generan un gran pico de tensión

muscular. Resumiendo, las fibras musculares se pueden clasificar en rápidas o lentas según su velocidad de contracción, distribuyéndose formando un patrón a lo largo de los fascículos (Helliwell, 1999; Pette y Staron, 2000; Ross y Wojciech, 2012; Junqueira y Carneiro, 2015), donde las muestra de las fibras musculares presentan una forma poligonal y tamaño homogéneo, rodeado por una malla fina de colágeno (Fig. 5), (Sáez y cols., 2013).

3.2. Desarrollo y crecimiento del músculo esquelético

Durante el desarrollo embrionario se generan las fibras que conformarán el músculo tras el nacimiento. A partir de entonces los músculos crecen aumentando la longitud y el grosor de las fibras pero manteniendo su número durante el resto de la vida del organismo. El crecimiento del músculo está relacionado con el aumento del número de núcleos como consecuencia de la división de las células satélites, que actúan como mioblastos, y de su fusión a las fibras musculares existentes. El crecimiento en longitud de las fibras musculares se debe al aumento del número de sarcómeros, este incremento se produce solamente en los extremos. En ellos, la membrana plasmática forma invaginaciones donde se anclan los miofilamentos y ahí es donde se disponen los nuevos miofilamentos prolongando el número de sarcómeros de cada miofibrilla. El crecimiento en grosor se debe a la síntesis de nuevos miofilamentos que se adosan lateralmente a los ya existentes. Cuando las miofibrillas se vuelven excesivamente gruesas se seccionan longitudinalmente, con lo que aumenta su número (Gomez-Alvarez, 2007).

4. La geometría como una herramienta para entender la biología de los tejidos

Muchos conceptos geométricos han sido ampliamente utilizados para intentar comprender como se establece la arquitectura de los tejidos naturales (Chichilnisky, 1986; Classen et al., 2005; Hayashi and Carthew, 2004; Honda, 1978; Rivier et al., 1995). La aplicación de la geometría computacional nos permite modelar tejidos naturales a partir de tejidos artificiales que emulen las características geométricas de los primeros. A lo largo de los años se han ido definiendo y desarrollando algunas propiedades geométricas inherentes a los tejidos empaquetados, definidos como un conjunto de células poligonales que rellenan el tejido sin dejar huecos y sin solaparse. Esta composición compacta del tejido hace que emerjan una serie de propiedades y leyes que describen la organización del tejido.

4.1. Propiedades geométricas en tejidos empaquetados

- ❖ La primera propiedad inherente a los tejidos empaquetados es el “Teorema de Euler” donde se dicta que dado un tejido con un gran número de células, el número medio de vecinos de cada célula tenderá a seis (Weaire, 1974; Weaire y Rivier, 1984).
- ❖ La segunda propiedad fue enunciada en 1928 por Frederic T. Lewis quien halló la existencia de una correlación entre la forma y el tamaño de las células en la piel del pepino (*Cucumis sativus*). Lewis había encontrado que la media del área de las células con el mismo número de vecinos crece de forma aproximadamente lineal con respecto a dicho número de vecinos. En otras palabras células con áreas pequeñas tenderán a tener menor número de vecinos que células con áreas mas grandes (Lewis, 1928; Chiu, 1995; Sangwoo y cols., 2014).
- ❖ La tercera propiedad es la ley de Aboav-Weaire que enuncia que las células con un alto número de lados tienden a tener células vecinas con pocos lados y viceversa (Chiu S. , 1995). La ley de Aboav-Weaire fue surgió del estudio de los mecanismos de crecimiento de los policristales. Con el tiempo se ha observado en otras estructuras celulares (Le Caër y Delannay, 1993; Boots 1980; Pignol y cols., 1993).

4.2. Patrones geométricos

Con el objetivo de crear tejidos artificiales empaquetados que emularan las propiedades geométricas descritas previamente, en las que se relaciona el tamaño de una célula con su forma, se han definido una serie de modelos de patrones geométricos. Estos patrones han sido diseñados para describir algunos de los tejidos celulares (células planas cultivadas, células de tejidos epiteliales, etc.) empleando la mínima información y obteniendo poca desviación en los resultados con los tejidos originales. También se han empleado modelos de tejidos en los que las propiedades biofísicas de las células son modificables para simular con exactitud el comportamiento de un tejido determinado en diferentes condiciones.

4.2.1. Teselaciones de Voronoi

Una de las técnicas en geometría computacional para generar tejidos artificiales es el diagrama de Voronoi. El diagrama de Voronoi rellena el plano o el espacio formando

polígonos convexos (llamados células de Voronoi (Rogers 1964) o dominio de Dirichlet (Loeb, 1976)) que no se solapan (Voronoi, 1908). Cada célula de Voronoi emerge de una semilla y está definida por la cualidad de que todos los puntos pertenecientes a la célula están más cerca de su correspondiente semilla que a cualquier otra semilla (Fig. 1D). Dependiendo del número de semillas empleadas y de donde se dispongan éstas en el plano se pueden obtener infinitos diagramas de Voronoi. En la naturaleza es fácil encontrar tejidos que emulan un diagrama de Voronoi, de aquí su uso ocasional para analizar tejidos naturales (Gibson y Gibson, 2009; Hocevar y cols., 2010; Honda, 1978; Zhu y cols., 2001). Honda en 1978 intentó describir patrones celulares presentes en la naturaleza mediante las células de Voronoi empleando el centro de masa de las células que quería replicar. Para ello, empleó distintos patrones naturales (como el patrón difuso en gelatina de gotas en la solución de ferrocianuro de potasio, las caras de un cristal de Zinc, el patrón que forman las grietas del esmalte de la cerámica, etc.) para obtener su equivalente patrón geométrico mediante las células de Voronoi y evaluar su desviación en un intento de describir los tejidos naturales (Loeb, 1976; Honda, 1978). Honda demostró que las células de Voronoi ofrecían la posibilidad de describir, con ciertos límites, las células de un tejido epitelial. Un tipo especial de diagrama de Voronoi es el CVT (del inglés “Centroidal Voronoi Tessellation”) donde la semilla con la que se genera la célula de Voronoi se corresponde con el centro de masa de dicha célula (Du y cols., 1999). Esto da lugar a polígonos muy regulares y a una alta proporción de hexágonos en el diagrama.

4.2.2. Modelo de vértices

Para ir más allá de la réplica de distintos tejidos biológicos e intentar entender los distintos procesos que pueden afectar a un tejido durante su desarrollo o como consecuencia de una enfermedad, es necesario buscar un modelo computacional que tenga en cuenta la interacción entre las células. Estos modelos presentan una naturaleza inherentemente compleja pero pueden generar información muy relevante sobre el sistema. Un método de simulación que ha sido utilizado ampliamente para investigar la interacción entre células es el modelo de vértices (del inglés, “vertex model”), (Honda y Eguchi, 1980; Brodland, 2004; Hardin y Walston, 2004; Farhadifar y cols., 2007; Fletcher y cols., 2014).

El modelo de vértices es una aproximación discreta que representa cada célula mediante un polígono. Estos polígonos están representados por vértices y por lados los cuales se compartirán entre células adyacentes. Los vértices se moverán en respuesta a las distintas fuerzas que regulen a las células que lo comparten. Es común

realizar dos simplificaciones en las que se considera que cada vértice está sumergido en un medio viscoso que aplica una fuerza de arrastre y que este movimiento inercial va desvaneciéndose con el tiempo. Una diferencia a lo largo de los distintos modelos de vértices existentes en la literatura, radica en la definición de esta fuerza que se aplica sobre los distintos vértices. Algunas de las fuerzas comúnmente modeladas son la tensión o la elasticidad. El avance en este campo llevó a desarrollar un modelo en el que son tres componentes los que contribuyen a la fuerza que se ejerce sobre cada vértice. Los dos primeros términos describen la tensión en la membrana celular y el tercer término describe la presión interna (o la tendencia de la célula a crecer), (Weliky y Oster, 1990; Weliky y cols., 1991; Brodland y cols., 2006; Brodland y cols., 2007). En algunos casos, más que identificar las fuerzas explícitamente, podría ser preferible considerar las fuerzas que surgen como resultado de la minimización de la energía. Por ejemplo, mientras que la anterior descripción de la fuerza codifica la respuesta a la compresión celular y estiramiento en términos de tensión y presión, una alternativa podría ser evaluar la energía almacenada tanto en la célula como en la membrana y resumir este valor para todas las células. La energía total almacenada está en función de la posiciones de los vértices de las células que corresponde con el trabajo necesario para deformar la red celular. Algunos de los estudios basados en la aproximación de la energía fueron llevados a cabo por Honda y colaboradores (Honda y Eguchi, 1980; Honda y cols., 1982; Honda, 1983), quienes investigaron como las células epiteliales sometidas a procesos de relajación y de intercambio de vecinos pueden minimizar su superficie manteniendo constante su volumen. Una de estas funciones de energía que regula el movimiento de los vértices ha sido ampliamente usada en el estudio del epitelio del disco de ala de *Drosophila* (Farhadifar y cols., 2007; Fletcher y cols., 2014). Esta función considera las condiciones de deformación de área de las células, los cambios en la tensión y adhesión entre las células y evalúa la capacidad de una célula de cambiar de forma.

Objetivos

Esta tesis enmarcada dentro de la Biología del Desarrollo, tiene como objetivo principal caracterizar y comprender tanto la morfogénesis como el mantenimiento de la organización celular de tejidos empaquetados mediante el análisis de imagen computerizado.

❖ **Entender y cuantificar los cambios morfogénéticos que ocurren durante el desarrollo de los tejidos proliferativos evaluando dos estadios muy similares.**

Empleamos técnicas de la biología de sistemas y de teoría de grafos con el objetivo de estudiar el efecto que tienen los mecanismos de proliferación en la organización. Para ello utilizamos tejidos epiteliales de *Drosophila* en los estadios de larva y prepupa.

❖ **Analizar las generalidades de los tejidos empaquetados en términos de su organización.** Está descrito en la literatura la existencia de una distribución estereotipada del número de lados de las células en tejidos proliferativos y cómo la división celular, la reordenación celular o ambos mecanismos presentes en la proliferación son responsables de esta ordenación poligonal. La hipótesis de partida es que esta distribución estereotipada de polígonos no es exclusivamente debida a la proliferación.

❖ **Profundizar en los fundamentos físicos que regulan la organización de las células para la conformación de los tejidos.** Los tejidos empaquetados están sujetos a una serie de restricciones físicas. Nuestro objetivo es investigar si esto dirige su organización y de qué forma lo hace.

Materiales y Métodos

1. Materiales

Hemos establecido una base de datos sólida con imágenes cuidadosamente seleccionadas para emplearla como herramienta para una evaluación efectiva de los distintos sistemas biológicos propuestos.

1. 1. Tejidos empaquetados proliferativos

Se usaron tejidos de *Drosophila* y de pollo (*Gallus gallus*) como modelos de tejidos proliferativos para comprender el mecanismo que regula la organización epitelial de tejidos empaquetados. Nuestra base de datos consistió en:

- ❖ 15 muestras provenientes de discos imaginales del ala de *Drosophila* en la mitad del tercer estadio de larva (dWL, del inglés “*Drosophila* Wing Larva”).
- ❖ 16 muestras de la misma región en un estadio temprano de la prepupa de *Drosophila* (dWP, del inglés “*Drosophila* Wing Prepupa”).
- ❖ 10 muestras del disco de ala de *Drosophila* en estadio de prepupa que presentan una reducción en la expresión gen de la cadena pesada de la miosina II (*zipper*) (dMWP, del inglés “*Drosophila* Mutant Wing Prepupa”).
- ❖ 16 muestras provenientes del neuroepitelio del pollo (cNT, del inglés “chicken Neural Tube epithelium”).
- ❖ 3 muestras del ojo de *Drosophila* en estadio de prepupa (EYE).

Las imágenes dWL (tercer estadio de larva) y dWP (prepupa) fueron obtenidas a las 96 horas y 120 horas del periodo de crecimiento respectivamente (temperatura de 25°C). La mutación que se produce en dMWP es debida a la reducción de la expresión de miosina II usando la combinación genética *C765-GAL4>UAS-*zip*-RNAi* (Escudero y cols., 2011). Las imágenes analizadas, excepto EYE, fueron capturadas y segmentadas según se describe en el apartado de métodos en (Escudero y cols., 2011). Las muestras provenientes de EYE fueron capturadas como se describe en (Escudero y cols., 2007) y segmentadas manualmente. El proceso de segmentación sirve para identificar unívocamente las células dentro de una imagen, lo cual nos facilita la extracción de información para poder realizar un correcto análisis del tejido.

1. 2. Tejidos empaquetados no proliferativos

Como modelos de tejidos empaquetados no proliferativos se emplearon biopsias de músculos de humanos. Capturamos una región de interés (ROI, del inglés “region of interest”) con resolución de 1100 x 1100 píxeles de imágenes de 3072 x 4080 píxeles. Las ROIs fueron seleccionadas evitando pequeños artefactos surgidos en las muestras provocadas por la manipulación y proceso de tinción. Nuestra base de datos de biopsias humanas consistió en:

- ❖ 29 imágenes provenientes de 12 biopsias diferentes de bíceps control de humanos adultos (BCA, del inglés “Biceps Control Adult”).
- ❖ 12 imágenes provenientes de 6 biopsias diferentes de bíceps con atrofia neurógena de humanos adultos (BNA, del inglés “Biceps Neurogenic Atrophy”).

Las imágenes procedentes de biopsias de humanos fueron tratadas bajo un informe de consentimiento usando un protocolo estandarizado (Dubowitz y cols., 2007) y segmentadas según se describe en (Sáez y cols., 2013).

2. Extracción de características de imágenes naturales

En cada imagen (natural, diagrama de Voronoi y simulaciones) seleccionamos un subconjunto de células válidas a la que le calculamos diferentes características. Las células válidas fueron identificadas como sigue: Células localizadas en el borde de la imagen las llamamos “BorderCells” (BC); los vecinos de BC fueron llamados “NeighbourBorderCell” (NBC). Las células válidas las llamamos “ValidCell” (VC) y pertenecieron a este grupo todas las células presentes en la imagen excluyendo BC y NBC.

$$VC = Cell_k, \forall k \notin BC, NBC \quad (1)$$

Se creó un grafo tomando los centroides de cada célula como nodo que se enlaza con los centroides de células adyacentes. Para construir el grafo, se siguió el método descrito en (Escudero y cols., 2011) realizando solo una modificación. En nuestro caso se ha usado un radio $r=4$ para definir el círculo utilizado en la identificación de los vecinos de cada célula.

Una vez identificadas las células VC y definido el grafo que da estructura al tejido celular, se procedió a la extracción de características. Definimos un conjunto de características puramente geométricas que usamos para caracterizar las células:

- ❖ **Área:** El área de una célula viene definida como el sumatorio de todos los píxeles asociados a dicha célula.

$$a = \sum_{j \in \Omega} 1 \quad (2)$$

Donde “a” es el área de la célula y “ Ω ” es el conjunto de píxeles de la célula.

- ❖ **Área Normalizada:** Normalizamos el área de cada célula del tejido a través del área media del conjunto de células presentes en el tejido. El uso del logaritmo hace que los valores se distribuyan de forma similar a la normal facilitando la comparación entre distintas clases de tejidos.

$$an_i = \log_{10} \left(\frac{a_i}{a} \right), \forall i \in [1, m] \quad (3)$$

Donde “ an_i ” es el valor del área normalizado de la célula “i”, “m” es el número de VC en la imagen, “ a_i ” es el área de la célula “i” y “ \bar{a} ” es la media de área de VC.

- ❖ **Eje mayor:** Longitud (en píxeles) del eje mayor de la elipse que tiene el mismo momento central de 2º orden normalizado que el de la célula bajo estudio.
- ❖ **Eje menor:** Longitud (en píxeles) del eje menor de la elipse que tiene el mismo momento central de 2º orden normalizado que el de la célula bajo estudio.
- ❖ **Envolvente convexa:** Definiendo la envolvente convexa como el polígono convexo más pequeño que puede contener a la célula. Se halla la proporción de píxeles de dicha envolvente convexa perteneciente también a la célula como:

$$r_i = \frac{a_i}{ac_i}, \forall i \in [1, m] \quad (4)$$

Donde “ r_i ” representa la proporción de área de la célula “i” (“ a_i ”) con respecto al área del polígono convexo que contiene a la célula (“ ac_i ”) y “m” es el número de VC en la imagen.

- ❖ **Media de vecinos de los vecinos de una célula:** Característica usada para determinar la ley de Aboav-Weire.
- ❖ **Relación de ejes:** Ratio entre el eje mayor y menor de cada célula.
- ❖ **Vecinos:** Número de células vecinas a la célula bajo estudio.

Definimos una serie de características que expresan relación geométrica de una célula con el valor medio de sus vecinos y un conjunto de características propias de grafos. En el estudio de grafos el grado de un nodo se conoce como el número de vecinos que tiene dicho nodo. Para extraer el conjunto de características fue necesario definir la matriz, W , de conexión de las células VC presentes en el tejido, la cual está caracterizada por ser una matriz simétrica al tratar con un grafo no dirigido. Cada posición w_{ij} de la matriz almacena la distancia en píxeles entre la célula “ i ” y “ j ”:

- ❖ **Asortatividad o emparejamiento:** La assortatividad se mide a través del coeficiente de correlación de Pearson de los grados de los nodos de un enlace (Newman, 2002).

$$r = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2} \quad (5)$$

Donde “ j_i ” y “ k_i ” son los grados de los nodos correspondiente al i -ésimo enlace con $i=1, \dots, M$. “ r ” toma valores comprendido en el rango $-1 \leq r \leq 1$. El emparejamiento mide la preferencia de los nodos de una red por unirse a otros que presentan un grado similar. Un “ r ” positivo refleja que los nodos tienden a unirse con grados iguales o similares. Un coeficiente de emparejamiento negativo refleja la conexión de nodos con alto grado con nodos de bajo grado.

- ❖ **Coeficiente de agrupación:** Nos define como de agrupados están entre sí los vecinos de un nodo. Si el nodo está agrupado con todos los nodos del grafo, tendrá un coeficiente de agrupación máximo. Un valor de coeficiente pequeño indicará un nodo poco agrupado en el grafo.

$$C_i(W) = \frac{(W^{1/3})_{ii}^3}{d_i(d_i - 1)} \quad (6)$$

$$C = \frac{1}{N} \sum_{i=1}^N C_i \quad (7)$$

Donde “ d_i ” es el grado del nodo i en la red. “ C_i ” representa el coeficiente de agrupación del nodo “ i ”. “ C ” representa el coeficiente de agrupación para todo el grafo (Fagiolo, 2007).

❖ **Conectividad algebraica:** Se encuentra definida como el segundo autovalor más pequeño del laplaciano. (Laplaciano: grado de la matriz menos la matriz adjunta. Matriz adjunta: Matriz con filas y columnas etiquetadas a través de los nodos del grafo, con 1 ó 0 en la posición “ w_{ij} ” según si “ i ” y “ j ” son adyacentes o no). La conectividad algebraica es dependiente tanto del número de nodos, como de la manera en que estos nodos están conectados. En un grafo aleatorio, la conectividad algebraica disminuye con el número de nodos y aumenta con la media de grados del nodo.

❖ **Correlación de Pearson:** La correlación de Pearson refleja el grado de linealidad entre los nodos de la red atendiendo al peso de los enlaces. Se empleó para el cálculo de esta característica el método desarrollado por el Dr. Daniel Whitney (Whitney, 2006).

❖ **Densidad:** Definida como la fracción de conexiones actuales entre las posibles conexiones. El peso de las conexiones fueron ignorados en estos cálculos.

$$D = \frac{K}{CP} \quad (8)$$

$$CP = \frac{N*(N-1)}{2} \quad (9)$$

“ K ” representa el número de conexiones actuales y “ CP ” el número de conexiones potenciales. “ N ” es el número de nodos en el grafo. Este parámetro nos da una medida de la densidad de enlaces en el grafo. La densidad toma valores en el rango $0 \leq D \leq 1$.

❖ **Diámetro:** Está definido como la máxima excentricidad entre todos los nodos del grafo.

- ❖ **Eficiencia:** El concepto de eficiencia nos mide como de eficiente intercambia información un grafo. Está relacionado por tanto con el concepto de coeficiente de agrupación. Está definida como la media de los valores inversos de la longitud del camino más corto.

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \quad (10)$$

Donde “ d_{ij} ” representa la distancia más corta entre el nodo “ i ” y el nodo “ j ” del grafo bajo estudio y “ N ” representa el número de grafos en la red (Latora y Marchiori, 2001).

- ❖ **Excentricidad:** La excentricidad de un nodo en un grafo está definida como la máxima distancia geodésica entre dicho nodo y cualquier otro nodo del grafo. Es decir, como de lejos está un nodo del nodo más distante de él en el grafo.
- ❖ **Fuerza:** La fuerza de un nodo se define como la suma de los pesos de los enlaces conectados al nodo, donde el peso de los enlaces, en nuestro caso, viene dado por la distancia, en píxeles, entre dos células.

$$S_i = \sum_{j \neq i} w_{ij} \quad (11)$$

- ❖ **Intermediación:** Es un indicador de la centralidad del nodo en la red. Fracción de todos los caminos cortos en grafo que contiene un determinado nodo. De tal forma, que nodos con altos niveles de intermediación participan en un alto número de caminos cortos. Se trata de una medida de centralidad de nodo en la red.
- ❖ **Modularidad:** Estadístico que cuantifica el grado por el cual el grafo puede ser subdividido en concentraciones de nodos o “clusters”. Grafos con alta modularidad presentan conexiones densa entre nodos del mismo “cluster” pero pocas conexiones entre nodos de distinto “cluster” (Good y cols., 2010).
- ❖ **Radio:** Está definido como la mínima excentricidad entre todos los nodos de la red.
- ❖ **S-metric:** Se define como la suma de productos de los grados a través de todos los enlaces. Se utilizó la definición propuesta por (Li y cols., 2005).

$$s = \sum_{(i,j) \in \mathcal{E}} d_i d_j \quad (12)$$

Donde “ d_i ” expresa el grado del nodo “ i ” (número de vecinos) y “ \mathcal{E} ” define el conjunto de enlaces presentes en el grafo. S-metric mide el alcance en el que el gráfico presenta concentraciones de nodos conectados. Se maximiza cuando nodos con un alto número de grados se unen con otros nodos con alto número de grados.

- ❖ **Transitividad:** Versión alternativa del coeficiente de agrupación. Se entiende como transitividad el porcentaje de triangulaciones de nodos que son transitivas. Si tenemos 3 nodos y el nodo “ A ” está enlazado con “ B ” y el nodo “ B ” está enlazado con “ C ”, entonces entendemos por triangulación transitiva si nodo “ A ” está enlazado con “ C ”. En otras palabras, nos da la probabilidad de que dos vecinos de un mismo nodo sean vecinos entre ellos. Al tratar con una red con pesos, la transitividad nos mide la capacidad de dicha red para pasar de un nodo a otro.

3. Evaluación de diferencias entre 2 categorías de imágenes

Hemos evaluado las diferencias existentes entre 2 grupos de imágenes a distintos niveles. Se desarrolló en primer lugar un método que nos permitió evaluar los grupos a nivel celular. Se empleó la distribución de número de vecinos para capturar información sobre la organización y poder ver como varía en los 2 grupos. Por último se evaluaron las diferencias entre 2 grupos de imágenes empleando el método PCA.

3.1. Comparación de células individuales: Correlación

Hemos desarrollado un análisis de correlación para evaluar la similitud entre las células pertenecientes a la imagen. Hemos definido cada célula empleando 14 características: “Media de área”, “Media de eje mayor”, “Media de eje menor”, “Media de relación de ejes”, “Media de envolvente convexa”, “Media de relación de área de vecinos”, “Media de relación de eje mayor de vecinos”, “Media de relación de eje menor de vecinos”, “Media de relación de relación de ejes de vecinos”, “Media de relación de envolvente convexa de vecinos”, “Media de relación de fuerza”, “Media de coeficiente de agrupación”, “Media de excentricidad” y “Media de intermediación”.

Una vez extraídos los 14 valores para cada una de las “ m ” células VC presentes en todos los tejidos, procedimos a realizar una comparación exhaustiva entre cada par de células. El proceso de correlación comparó pares de células a lo largo de las 14 características seleccionadas. Los datos fueron definidos como:

$$\begin{aligned}
cell\ 1 &= [característica\ 1, característica\ 2, \dots, característica\ 14] \\
cell\ 2 &= [característica\ 1, característica\ 2, \dots, característica\ 14] \\
&\dots \\
cell\ m &= [característica\ 1, característica\ 2, \dots, característica\ 14]
\end{aligned} \tag{13}$$

La siguiente ecuación nos proporcionó el valor de correlación para cada par de células.

$$Corr = \frac{\sum_{n=1}^{14} [cell\ 1(n) - \overline{cell\ 1}] * [cell\ 2(n) - \overline{cell\ 2}]}{\sigma_{cell\ 1} * \sigma_{cell\ 2}} \tag{14}$$

Donde " $\sigma_{cell\ 1}$ " representa la desviación estándar de los datos provenientes de " $cell\ 1$ " y " $\overline{cell\ 1}$ " es la media de los valores normalizados de las 14 características de " $cell\ 1$ ". El valor " $Corr$ " obtenido debía estar en el rango entre 0 y 1; siendo los valores más cercanos a 1 indicativo de una alta similitud entre las células comparadas. Esta medida fue realizada para cada pareja de células. Definimos un umbral para obtener pares de células que presentasen una similitud clara respecto al resto. El umbral fue aplicado para determinar la variable " $distancia$ ", la cual utilizamos para determinar si un par de células eran muy similares ($distancia=1$) o si presentaban alguna diferencia ($distancia=0$):

$$\begin{aligned}
distancia\ (cell\ 1, cell\ 2) &= 1 \quad si \quad Corr\ (cell\ 1, cell\ 2) \geq umbral \\
distancia\ (cell\ 1, cell\ 2) &= 0 \quad si \quad Corr\ (cell\ 1, cell\ 2) < umbral
\end{aligned} \tag{15}$$

El software libre Visone (descargado en <http://visone.info/html/download.html>) se empleó para visualizar la relación entre células. Este programa utilizó la información guardada en la variable distancia para construir la red de correlación de las células conectadas, es decir, una red donde 2 células eran conectadas si su valor de correlación se encuentra por encima del umbral.

Para determinar un valor de umbral adecuado fue preciso evaluar la red de correlación creada. Si el umbral era demasiado alto, la red presentaría pocas células relacionadas. Por otro lado, si el umbral era muy bajo, apenas podríamos obtener información de la enorme red resultante. Se examinaron diferentes umbrales para corroborar que nuestros resultados no dependían de la opción elegida. Se definió como umbral óptimo, aquel con el que se obtenía una red principal con 1500-2000 células. Visone puede manejar diferentes etiquetas que facilitan el análisis de la distribución de las diferentes características que definen a las células. Esto se usó para analizar las causas de la correlación.

3.2. Comparación de histogramas de frecuencia de polígonos: MANOVA

Para evaluar si la distribución de la característica número de vecinos de 2 grupos de imágenes era significativamente diferente se utilizó un análisis multivariado de varianza (MANOVA, del inglés “Multivariate Analysis of Variance”). El test MANOVA realiza una comparación entre 2 o más grupos independientes (en nuestro caso sólo 2) para ver si hay diferencias significativas para múltiples variables dependientes. Nuestra variable dependiente fue la característica porcentaje de células con N vecinos. Se ejecutó el test MANOVA comparando las células con 4, 5, 6, 7 y 8 lados. Se descartaron células con 3, 9 y 10 vecinos debido a que células de estos tipos no estuvieron presentes en todas las imágenes. Un *p value* menor que 0.05, indicaría que la distribución es significativamente diferente a lo largo de todas las variables dependientes.

3.3. Comparación de la organización tisular entre dWL y dWP

Este método lo realizamos para poder comparar dos categorías de tejidos tanto a nivel geométrico como topológico. Primero obtuvimos las 40 características, mostradas en la Tabla 1 y definidas en la sección 2 del presente capítulo, de las imágenes que conforman las categorías dWL y dWP. Los valores de las primeras 28 características se analizaron extrayendo la información de cada célula y calculando la media y la desviación estándar (D.E.) de todas las VC pertenecientes a la ROI de cada imagen. Las 12 características restantes se extrajeron directamente del grafo formado por los nodos pertenecientes a la ROI. Posteriormente, normalizamos los valores de las diferentes características para posibilitar la comparación entre éstas.

Se empleó el método basado en el análisis de componentes principales (PCA, del inglés “Principal Component Analysis”), (Fukunaga, 1990), el cual analiza las diferencias entre 2 o mas categorías usando un conjunto de características extraídas de las muestras de los tejidos. PCA transforma el conjunto de datos correlacionados del vector en un pequeño número de variables no correlacionadas llamadas componentes principales. La proyección maximiza la dispersión de los datos de un modo no sesgado. Esto permite la identificación de un conjunto de datos naturalmente separados (imágenes en nuestro caso). Estos datos se pueden visualizar gráficamente basándonos en su posición sobre el grafo PCA cuando se representan los 2 primeros componentes principales. Una vez que obtuvimos el grafo de PCA, usamos una variante del descriptor “Calinski-Harabasz” para evaluar de manera cualitativa el grado de separación entre las 2 categorías de imágenes (Costa y Cesar, 2009).

$$W = \sum_{i=1}^C \sum_{j=1}^{M_i} (x_i(j) - \bar{x}_i)(x_i(j) - \bar{x}_i)^T \quad (16)$$

$$B = \sum_{i=1}^C M_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T \quad (17)$$

$$T = W + B = \sum_{j=1}^M (x(j) - \bar{x})(x(j) - \bar{x})^T \quad (18)$$

$$\text{Descriptor PCA} = \text{trace} \left(\frac{B}{W} \right) \quad (19)$$

Donde en “ x ” se almacenan las “ M ” coordenados de todos los tejidos representados en el grafo de PCA, “ x_i ” es el subconjunto de “ M_i ” coordenadas de “ x ” que pertenecen al conjunto de imágenes de la categoría “ i ”, “ $x_i(j)$ ” indica el centroide j -ésimo asignado al conjunto de imágenes de la categoría “ i ”, “ \bar{x}_i ” indica la distancia media de todas las coordenadas pertenecientes al conjunto de imágenes de la categoría “ i ”, es decir, el centroide del conjunto de imágenes de la categoría “ i ”, “ \bar{x} ” indica la distancia media de todas las coordenadas del conjunto “ x ” indicando el centroide del conjunto total de datos. “ W ” (matriz intragrupo) nos proporciona información de la dispersión de los datos de un conjunto de tejidos pertenecientes a una categoría mientras que “ B ” (matriz intergrupo) nos informa cómo de separadas están las categorías. Por tanto, las matrices intergrupo e intragrupo resumen la dispersión del conjunto de datos (Fukunaga, 1990; Costa y Cesar, 2009). Como consecuencia, los datos pertenecientes a una categoría que estén distribuidos de forma compacta y estén separados de los tejidos de otras categorías tendrán valores pequeños en “ W ” y valores grandes en “ B ”. Es por ello, que a mejor separación de datos mayor será el ratio entre “ B ” y “ W ”.

Debido al alto número de características extraídas en comparación con el bajo número de muestras disponibles para el estudio, se disminuyó previamente el número de características para efectuar la comparación entre las dos categorías de tejidos. Para seleccionar las características relevantes se aplicó el PCA usando todas las posibles combinaciones de 2 características. Mantuvimos las 10 combinaciones de 2 características con el valor del descriptor del PCA más alto. Se aplicó otra vez el PCA combinando cada una de las características con las 10 mejores combinaciones de 2 características seleccionadas en el paso previo y mantuvimos las 5 combinaciones con el valor más alto del descriptor del PCA para cada una de las 10 parejas. Así el programa preseleccionó 50 tríos de características. Se repitió el mismo proceso una vez más, pero añadiendo en esta ocasión las 2 mejores características por combinación para realizar las comparaciones con PCA, acumulando de este modo 100 cuartetos de características. A partir de entonces se ejecutó el proceso iterativamente

añadiendo solo una característica por iteración. El proceso se paró cuándo se seleccionaron 7 características o cuando el valor para el descriptor del PCA resultó ser menor que el correspondiente a la iteración previa. La combinación de características seleccionadas fue aquella combinación que presentaba el mayor descriptor de PCA.

Para entender el porqué del uso de este proceso iterativo se debe tener en cuenta que cuando se selecciona un conjunto de características, existen algunas que por sí solas son muy pobres en la clasificación, pero que unidas a otras la capacidad de clasificación aumenta, incluso en el caso de que 2 características fueran irrelevantes, al unir las, pueden dar resultados mejores que la adición de una característica individual. Es por ello que aplicando este proceso iterativo a nuestro método, nos aseguramos que estamos ante la mejor combinación de características que consigue una mejor identificación entre 2 categorías distintas de tejidos.

3.4. Aleatorización de imágenes

El número de combinaciones posibles para formar 2 grupos homogéneos sin repetición viene dado por la expresión:

$$C_{m,n} = \frac{m!}{n!(m-n)!} \quad (20)$$

En nuestra base de datos disponíamos de 2 grupos en los que se tenían 15 tejidos provenientes de dWL y 16 tejidos provenientes de dWP, teniendo un total de 31 muestras. El número de combinaciones de estos 31 tejidos para formar 2 nuevos grupos con 15 y 16 tejidos sería $C_{15,16} = 300540195$. El número de combinaciones totales resultó ser muy elevado y aunque tratáramos con solo una fracción de estas combinaciones, para movernos en el entorno de la probabilidad, el número seguiría siendo aún intratable a efectos prácticos. Es por ello, que para realizar la creación de 2 nuevos grupos empleando diferentes combinaciones de los 31 tejidos disponibles, se siguió el criterio de tratar solo aquellas combinaciones en que los grupos estuvieran totalmente mezclados.

Esto se consiguió eligiendo al azar la mitad de las muestras de una categoría y la mitad de las muestras de la otra categoría para crear un nuevo grupo artificial “mixA”. Las imágenes restantes formaron el grupo “mixB”. El número posible de combinaciones para formar los 2 nuevos grupos de imágenes siguiendo esta estrategia se redujo a $C_{16,8} = 12870$. Para tener un número representativo de combinaciones y atendiendo al número estimado de combinaciones sin repetición se optó por generar finalmente 4000 combinaciones aleatorias.

Se compararon mediante PCA cada uno de los 2 nuevos grupos “mixA” y “mixB” que se crearon en cada una de las 4000 aleatorizaciones de imágenes, obteniendo el descriptor de PCA correspondiente que define cómo de diferentes son estos 2 grupos. Representamos el histograma de los 4000 descriptores de PCA resultado de comparar los grupos realizados. Se da el sistema por válido si el grueso de las comparaciones de grupos aleatorios presenta una peor clasificación que la combinación de las categorías originales. Si ocurriese esto se evidenciaría la existencia de un mecanismo biológico de trasfondo con el que se consigue separar 2 categorías de tejidos muy similares, siendo esta separación no casual debido a que no se consiguen los mismos resultados en un porcentaje alto de combinaciones aleatorias.

4. Simulación de tejidos naturales empaquetados

4.1. Diagramas de Voronoi y vía CVT

El diagrama de Voronoi es un procedimiento geométrico que divide el espacio en un número determinado de regiones o células. Para crear los diagramas de Voronoi utilizados en este estudio se especificaron al azar un conjunto de “ n ” semillas en un plano $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$. Para cada semilla (x_k, y_k) se generó una célula Voronoi (“ $Cell_k$ ”), la cual fue formada asignándole todos los píxeles que estuvieran más cercanos a la semilla “ k ” que a cualquier otra. De este modo, un píxel (a, b) pertenecía a $Cell_k$ cuando la distancia a la semilla (x_k, y_k) era menor o igual que la distancia a cualquier otra semilla.

$$L_k = \sqrt{(x_k - a)^2 + (y_k - b)^2}, \forall k \in [1, n] \quad (21)$$

$$\min(L_k) \Rightarrow (a, b) \in Cell_k \quad (22)$$

Para la obtención de diagramas de Voronoi que presentasen una composición de células más homogéneas se utilizó el algoritmo de Lloyd. El procedimiento consistió en el cálculo del centroide (“ c_k ”) de cada célula $Cell_k$ para usarlo como semilla (x_k, y_k) de un nuevo diagrama. Este procedimiento creó un diagrama más homogéneo en términos de forma y tamaño de las células. Las iteraciones sucesivas de este proceso produjeron un conjunto de diagramas de Voronoi que conformaron la denominada vía CVT (del inglés “Centroidal Voronoi Tessellation”).

En nuestros experimentos, para obtener imágenes de tejidos empaquetados artificiales, usamos 500 puntos aleatorios a modo de semillas sobre una imagen de 1024x1024 píxeles y ejecutamos el procedimiento de Voronoi, obteniendo diagrama 1 o Poisson-Voronoi. Al estar tratando con un procedimiento aleatorio fue necesario trabajar con un determinado número de muestras de cada diagrama de la vía CVT. Hemos utilizado 20 inicializaciones diferentes de Poisson-Voronoi para obtener datos muestrales de la vía CVT (20 realizaciones independientes).

4.1.1. Modelo continuo de la vía CVT y densidad de probabilidad

Para transformar los diagramas obtenidos de forma discreta que conformaban la vía CVT (diagramas 1 a 200) a una escala continua, tomamos como referencia el porcentaje de hexágonos (que va incrementando en los sucesivos diagramas de la vía). En los diagramas de Voronoi, conociendo la media del porcentaje de hexágonos de las 20 realizaciones de imágenes se puede obtener unívocamente el porcentaje del resto de polígonos. Extrajimos para todos los diagramas de la vía CVT representados en la Tabla 4 (i.e. 20 realizaciones de D1, D2, D3, D4, D5, D6, D10, D20, D30, D40, D50, D100, D200) la dupla (P_6 , P_x) para estimar el porcentaje de un cierto polígono donde " P_6 " representaba el porcentaje de hexágonos y " P_x " indicaba el porcentaje de polígonos con " x " lados (siendo " x " igual a 4, 5, 7 o 8). No se incluyeron el resto de polígonos debido a que aparecen con muy poca frecuencia (siempre menor que el 5%, y 0% en la mayoría de los diagramas de Voronoi, Tabla 4). Aplicando un ajuste de curvas, diseñamos una serie de funciones matemáticas polinómicas que mejor se ajustaban al conjunto de datos (Anexo I), estando el rango de ajuste del porcentaje de hexágonos comprendido entre [25-75]. Obtuvimos 20 funciones, una por cada realización de CVT para cada tipo de polígonos que quisimos evaluar. El rango [25-75] comprende el conjunto de los porcentajes de hexágonos en los diferentes diagramas del CVT. Para representar la vía CVT seleccionamos 500 números aleatorios en un rango de 25 a 70 para cada función que resuelve " P_x " (este rango fue ligeramente diferente al usado en el experimento de ajuste de curvas debido a que permite una mejor visualización de las imágenes naturales). Los 10000 puntos resultantes nos proporcionaron la información de densidad de probabilidad. Esta información se representó en escala de grises donde las zonas negras y blancas representaban regiones con mayor y menor probabilidad respectivamente. Usamos este gráfico para visualizar la similitud en términos de la distribución de polígonos de las imágenes naturales y las simulaciones con la vía CVT.

4.1.2. Diferencias estadística entre la vía CVT y las imágenes naturales

Una vez que obtuvimos el modelo continuo de la vía CVT, lo empleamos para cuantificar las diferencias entre las imágenes naturales y los diagramas de Voronoi. Seleccionamos el porcentaje medio de hexágonos, " P_6 ", de cada conjunto de imágenes naturales e introdujimos el valor en las 20 ecuaciones continuas de la vía CVT que habíamos generado, para extraer los 20 valores de porcentaje de cuadrados, pentágonos, heptágonos y octógonos que "idealmente" se corresponde con cada conjunto de imágenes. Comparamos los valores reales de " P_x " de las imágenes naturales y el correspondiente valor "ideal" en el modelo continuo de la vía CVT. En primer lugar evaluamos si las muestras "reales" e "ideales" presentaban una distribución similar y una varianza parecida empleando para ello los test Kolmogorov-Smirnov y F-Snedecor respectivamente. En los casos en que los datos presentaron la misma distribución pero no la misma varianza, utilizamos el test de Welch de dos colas para comparar las medias de ambos grupos. En los casos en que los datos presentaron diferencias en la distribución, utilizamos el test de Wilcoxon para comparar las medias de ambos grupos. Empleamos el test de t-student con 2 colas pareadas para comparar las medias en los casos donde tanto la distribución como la varianza de los dos conjuntos de datos eran similares (Tabla 5). Un *p value* menor que 0.05 indicaba que el valor real de " P_x " de las muestras biológicas era significativamente diferente del valor ideal de " P_x " del modelo continuo de la vía CVT. Consideramos únicamente relevante para el estudio la comparación de " P_x " con una representación de muestras mayor del 5% del total de células (los porcentajes más bajos a menudo provocaban una gran variabilidad incluso en los diagramas de Voronoi).

4.2. Simulación de tejidos biológicos por ordenador

Hemos utilizado simulaciones de tejidos biológicos por ordenador basados en un modelo de vértices que ha sido descrito previamente en otros trabajos (Farhadifar y cols., 2007; Mao y cols., 2011). En este modelo las células fueron tratadas como polígonos. La posición de los vértices de estos polígonos delimitaban la forma celular. La conexión entre pares de vértices representaba el contorno compartido entre 2 células vecinas. La energía del tejido completo (ecuación 23) fue calculada como la contribución energética de la desviación de cada célula de su tamaño ideal (*a*), la suma de tensión y adhesión en cada unión de células (*b*) y sobretodo de la contractibilidad

de cada célula (c), reflejando la tendencia básica de la actomiosina del citoesqueleto para resistir la deformación de una forma isométrica.

$$E(R_i) = \sum_{\alpha} \frac{K_{\alpha}}{2} (A_{\alpha} - A_{\alpha}^{(0)})^2 + \sum_{\langle i,j \rangle} \Lambda_{ij} l_{ij} + \sum_{\alpha} \frac{\Gamma_{\alpha}}{2} L_{\alpha}^2 \quad (23)$$

(a) (b) (c)

De tal forma que, “ $E(R_i)$ ” es la energía total para las posiciones de los vértices “ R_i ”, definiendo “ N_C ” células ($\alpha=1 \dots N_C$), compuestas de “ N_V ” vértices ($i=1 \dots N_V$). En el cálculo de la “elasticidad del área” (a), “ K_{α} ” es el coeficiente de elasticidad, “ A_{α} ” es el área actual de la célula y “ $A_{\alpha}^{(0)}$ ” es el área ideal. En el cálculo de la “tensión lineal” (b), el sumatorio trabaja sobre todos los vértices conectados $\{i, j\}$, “ Λ_{ij} ” es el coeficiente de tensión lineal y “ l_{ij} ” la longitud de la unión entre los vértices. En el cálculo de la “contractibilidad cortical” (c), “ L_{α} ” es el perímetro de la célula y “ Γ_{α} ” es el coeficiente de contractibilidad. Para consultar detalles sobre la relajación de los tejidos en las simulaciones consultar (Farhadifar y cols., 2007; Mao y cols., 2011).

4.2.1 Simulación de tejidos no proliferativos

Para las simulaciones no proliferativas, primero generamos 20 diagramas de Voronoi de forma aleatoria de 500 células y se relajan aplicando el algoritmo de Lloyd’s durante 4 iteraciones. Estas iteraciones proporcionaron 20 imágenes de partida equivalentes al diagrama 5 de la vía CVT. Para la simulación de células saludables o control, se usaron los parámetros ajustados en estudios previos (Mao y cols., 2013), donde el área ideal de la célula, $A_{\alpha}^{(0)}$, era 1.0; el coeficiente de elasticidad, K_{α} , era 1.0 (el cual fue invariante en todas las simulaciones); la tensión lineal, Λ_{ij} , era 0.26; y el coeficiente de contractibilidad, Γ_{α} , era 0.0.

Para las simulaciones en las que se intentaba mimetizar el comportamiento de tejidos que presentaban atrofas neurógenas, se generaron aleatoriamente células afectadas de tal forma que comprendieran el 10 por ciento de la población total. Las células afectadas se generaron reduciendo el área ideal, $A_{\alpha}^{(0)} = 0.3$; reduciendo la tensión lineal, $\Lambda_{ij} = -0.104$ (valor negativo indica que domina la adhesión en las uniones intercelulares, provocando una tendencia a incrementar la zona de contacto); e incrementando levemente la contractibilidad total, $\Gamma_{\alpha} = 0.05$. Como control, se desarrolló un procedimiento similar tomando como valor normal del área ideal $A_{\alpha}^{(0)} = 1.0$ y manteniendo la reducción de la tensión lineal, $\Lambda_{ij} = -0.104$ y $\Gamma_{\alpha} = 0.05$. Todas las simulaciones fueron desarrolladas en un tiempo equivalente a 240 segundos, el

análisis de estas imágenes fue llevado a cabo por el mismo método de análisis de imágenes utilizado para imágenes naturales.

4.2.2 Simulación de tejidos proliferativos

Las simulaciones que investigaron los efectos de la división celular fueron procesadas durante un tiempo equivalente de 18 horas, con un ciclo medio de división celular de cinco horas, donde las cuatro primeras horas de crecimiento fueron seguidas de un proceso de división en el tiempo restante. Durante la fase de crecimiento, el área ideal ($A_{\alpha}^{(0)}$) de la célula fue incrementada gradualmente a $2x A_{\alpha}^{(0)}$. En los escenarios de control, las células tenían permitido dividirse una vez que sus áreas alcanzasen 1.95 veces el área que tenían al comienzo del crecimiento. Cuando se producía la división, el área ideal de las células hijas era establecida a $A_{\alpha}^{(0)}$, y se iniciaba la fase de reposo. El establecimiento de los parámetros de los tejidos saludables son los mismos que los enunciados anteriormente en la simulación de tejidos no proliferativos ($A_{\alpha}^{(0)} = 1.0$, $K_{\alpha} = 1.0$, $\Lambda_{ij} = 0.26$ y $\Gamma_{\alpha} = 0.0$).

Para intentar reproducir la reducción de la proteína miosina II en tejidos proliferativos se realizaron tres procedimientos. En primer lugar, conociendo que la desregularización de esta proteína afecta directamente a la tensión superficial de la célula, se asignó un parámetro de tensión reducido a células seleccionadas aleatoriamente siguiendo una distribución uniforme en el rango $\Lambda_{ij} = 0.05 - 0.2$. Para los casos restantes se aplicó un criterio adicional conociendo que la pérdida de tensión en la superficie celular afecta al proceso activo de división celular. Establecimos un umbral, de tal forma que si la tensión celular era mayor que este umbral se completaría la división celular satisfactoriamente; de lo contrario estas células crecerían en tamaño, sin poder dividirse. Una célula sin la capacidad de dividirse se quedaría estancada en la fase mitótica y no comenzaría una segunda vuelta de división celular, manteniendo su $A_{\alpha}^{(0)}=2$. Se realizaron simulaciones utilizando 2 umbrales, 30 por ciento de la tensión control ($\Lambda_{ij} = 0.078$) y 40% de la tensión control ($\Lambda_{ij} = 0.104$).

5. Análisis estadístico de la distribución del área celular en tejidos naturales y diagramas de Voronoi

Para evaluar el comportamiento geométrico que presenta un tejido determinado se evaluó estadísticamente la distribución que presenta la característica área normalizada extraída para todas las VC. Generamos el histograma utilizando un paso muestral de

0.02 unidades. Determinamos algunos estadísticos para describir y comparar la forma que presenta la distribución de área en las distintas categorías de tejidos estudiadas:

- ❖ **Media:** Tendencia central de la distribución de probabilidad.
- ❖ **Desviación estándar:** Mide como se dispersan los datos respecto a la media.
- ❖ **Moda:** Valor numérico que aparece con más frecuencia en el conjunto de datos.
- ❖ **Mediana:** Valor numérico que separa la mitad más alta de un conjunto de datos de la mitad más pequeña.
- ❖ **Skewness:** Mide la asimetría que presenta la distribución de probabilidad respecto a la media. Un sesgo negativo señala que la cola en el lado izquierdo de la función densidad de probabilidad es más larga que la cola en el lado derecho. Por otro lado, un sesgo positivo nos indica que la cola en el lado derecho es más larga que en el lado izquierdo.
- ❖ **Kurtosis:** Kurtosis es una medida de cómo de atípica es una distribución. La kurtosis de una distribución normal es 3. Las distribuciones que tengan una mayor concentración de datos en torno a la media que la normal, tendrán una kurtosis mayor de 3. Las distribuciones que tengan una menor concentración de datos en torno a la media presentarán una kurtosis menor de 3.
- ❖ Adicionalmente, obtuvimos los percentiles del Área.

6. Selección de subconjuntos de células

6.1. Células con atrofia neurógena

Para identificar células atróficas en imágenes de BNA seguimos las indicaciones y consejos de 2 especialistas en enfermedades neuromusculares (patólogo y neuróloga del Hospital Virgen del Rocío, Sevilla, España). Señalamos como células atróficas aquellas que presentan 2 condiciones: ser alargadas y pequeñas. Estas condiciones fueron tomadas usando el siguiente procedimiento: Para seleccionar células alargadas empleamos la característica relación de ejes (eje mayor/eje menor) que nos da una medida de la redondez de la célula.

Seleccionamos como células atróficas en las imágenes generadas por simulación aquellas células válidas que tienen una “relación de ejes” mayor que 3 y un “valor de área normalizada” menor que -0.2. En las imágenes Control generadas por simulación seleccionamos el 10% de células más pequeñas dentro del conjunto de células válidas atendiendo a la característica de área normalizada. En las imágenes BNA se toma el 17.5 % de células más pequeñas. Por último se emplea el 10 % de células más pequeñas según la característica de área normalizada para realizar el análisis de imágenes BCA.

Resultados

1. Estudio de la influencia de la proliferación en la organización de tejidos epiteliales

Como objetivo principal de este capítulo, se pretende cuantificar y comparar la organización celular en distintos tejidos empaquetados. Analizaremos imágenes de un mismo tejido en dos estadios de desarrollo muy cercanos, donde el crecimiento del tejido está ligado a cambios de forma y organización celular.

1.1. Análisis de características celulares en muestras de larva y prepupa en el disco imaginal de ala de *Drosophila*

Hemos utilizado el epitelio del disco imaginal de ala de *Drosophila* como modelo para comprender los mecanismos que regulan la organización del tejido proliferativo. Específicamente, se han empleado muestras del tercer instar del disco de ala en estadio de larva (dWL) y en estadio de prepupa (dWP), (Materiales y Métodos 1.1). Estos 2 estadios proliferativos están separados solo por 24 horas de desarrollo. Durante este corto periodo se producen pocos cambios a nivel de superficie apical de las células. Por ello, las muestras eran muy parecidas visualmente (Fig. 6A). Se realizó un proceso de segmentación para simplificar la representación de las células de una imagen a otra más fácil de analizar que nos posibilite la rápida extracción de información (Escudero y cols., 2011). El proceso de segmentación de células en un tejido consiste en la identificación en una imagen de grupos de píxeles que comparten una característica en común, como por ejemplo, el color, la intensidad o la textura. Estos grupos de píxeles se identifican como un único objeto o célula mediante una etiqueta. Una imagen segmentada de un tejido muestra el conjunto total de células identificadas y su localización en la imagen original.

Usando las imágenes segmentadas, primero investigamos si podíamos encontrar diferencias a nivel de las células individuales de dWL y dWP. Para ello, definimos una serie de 14 características que podrían reflejar diferencias a nivel celular entre los 2 tipos de imágenes (Tabla 1, Materiales y Métodos 2). Las 5 primeras características se corresponden con medidas geométricas, mientras que el resto se corresponden con medidas topológicas que capturan información sobre diferentes aspectos que relacionan una célula con sus vecinos.

Una vez que extrajimos los valores de estas 14 características para cada célula, realizamos un análisis de correlación usando el conjunto de todas las células de ambos estadios (Tabla 2). Como resultado, obtuvimos una matriz que nos indicaba el grado de similitud entre cada pareja de células. Para facilitar el análisis de estos datos, se

representó la matriz como un grafo de células correlacionadas. Cada nodo del grafo hace referencia a una célula, que se conectó a las células que presentaron un coeficiente de correlación superior a un cierto umbral. El alto grado de similitud entre la mayoría de las células motivó el uso de un umbral elevado. Esto propició que células ligeramente diferentes quedaran desconectadas del grafo principal (el de mayor número de células) creando diversos grafos secundarios. El umbral se estableció arbitrariamente para disponer un grafo principal que nos proporcione suficiente información sobre la similitud entre células (Materiales y Métodos 3.1). Seleccionamos un umbral de 0.9975 que nos proporcionó un grafo principal con 1729 células conectadas (Fig. 6B,C).

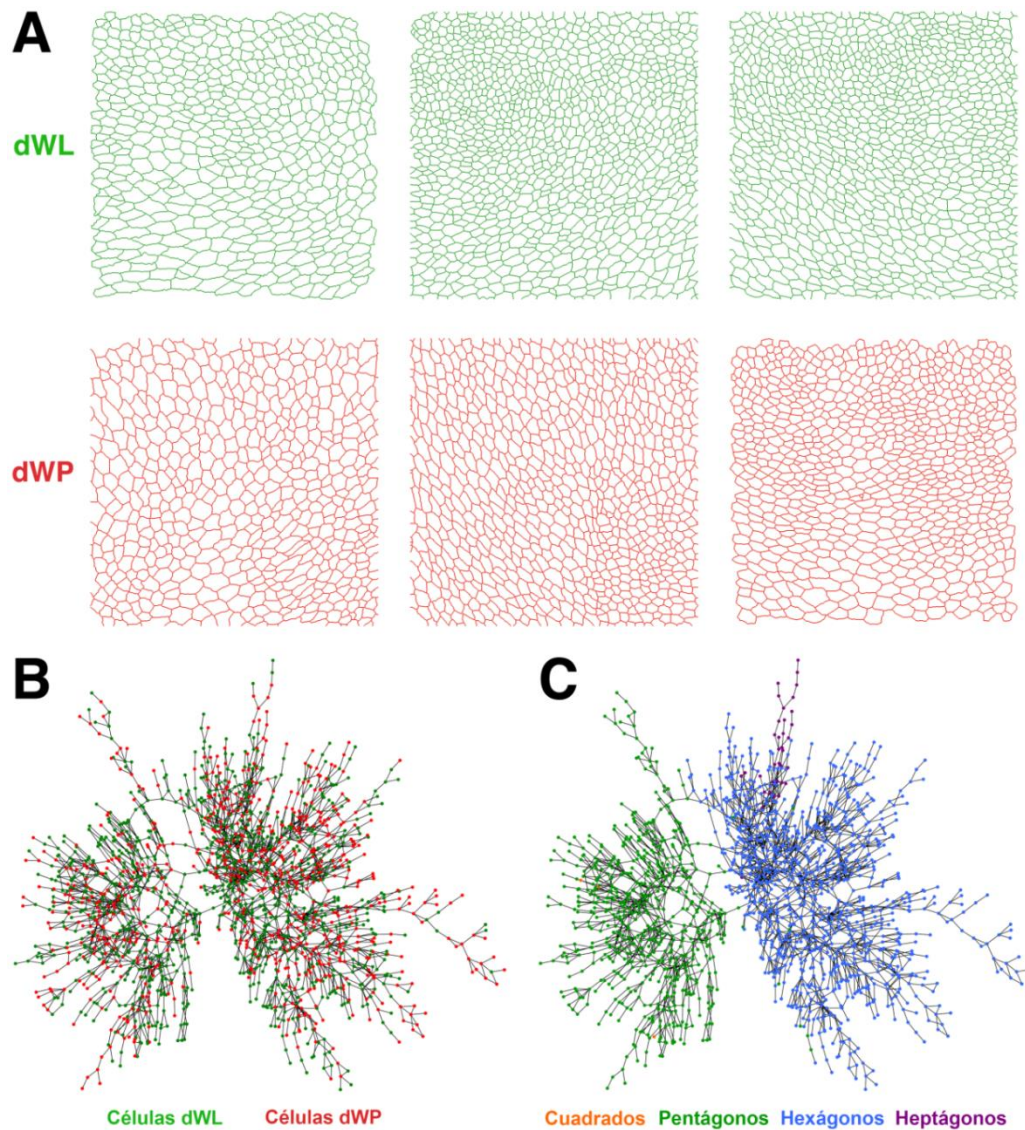


Fig. 6. Imágenes epiteliales y correlación entre células. **A)** Representaciones a partir de imágenes segmentadas de dWL (verde) y dWP (rojo) evidenciando una alta similitud entre ellas. **B)** Representación de grafos de células correlacionadas. Células provenientes de categoría dWL están marcadas en verde y células provenientes de dWP están marcadas en rojo. Los grafos contienen células de ambos tipos de imágenes donde cada célula está representada por un nodo. Los nodos se conectan si presentan una similitud más alta que un umbral de 0.9975. **C)** Representación del mismo grafo de células correlacionadas de la figura 1B, mostrando la distribución de lados de cada célula. Colores naranja, verde, azul y morado marcan células de 4, 5, 6 y 7 lados respectivamente. La imagen muestra la alta tendencia de las células a unirse con otras células con su mismo número de lados.

Características (CC.)		
CC. epitelial	Nombre	CC. celular
1	Media Área	1
2	D.E. Área	
3	Media Eje mayor	2
4	Media Eje menor	3
5	Media Relación de ejes	4
6	D.E. Relación de ejes	
7	Media Envolverte convexa	5
8	D.E. Envolverte convexa	
9	Media Vecinos	
10	D.E. Vecinos	
11	Media Relación de área de vecinos	6
12	D.E. Relación de área de vecinos	
13	Media Relación de Eje mayor de vecinos	7
14	D.E. Relación de Eje mayor de vecinos	
15	Media Relación de Eje menor de vecinos	8
16	D.E. Relación de Eje menor de vecinos	
17	Media Relación de Relación de ejes de vecinos	9
18	D.E. Relación de Relación de ejes de vecinos	
19	Media de Relación de Envolverte convexa de vecinos	10
20	D.E. de Relación de Envolverte convexa de vecinos	
21	Media de Fuerza	11
22	D.E. de Fuerza	
23	Media Coeficiente de agrupación	12
24	D.E. Coeficiente de agrupación	
25	Media Excentricidad	13
26	D.E. Excentricidad	
27	Media Intermediación	14
28	D.E. Intermediación	
29	Media longitud del camino más corto	
30	D.E. longitud del camino más corto	
31	Radio	
32	Diámetro	
33	Eficiencia	
34	Correlación de Pearson	
35	Conectividad algebraica	
36	S-metric	
37	Emparejamiento	
38	Densidad	
39	Transitividad	
40	Modularidad	

Tabla 1. Lista de características analizadas. La tabla muestra las 40 características empleadas en la comparación de imágenes epiteliales. Las 40 características se pueden clasificar en tres tipos: geométricas, relacionadas con la forma y el tamaño (1-8), características de red de la célula (9-28) y características de red del tejido (29-40). Las características de red recogen información de la organización. La tabla también señala las 14 características empleadas a nivel celular usadas en el ensayo de correlación. D.E. (Desviación Estándar).

Si estas 14 características eran útiles para distinguir las células de los dos estadios, deberíamos encontrar una agrupación con mayoría de células de dWL y otra agrupación con células de dWP. Este no fue el caso. El grafo contenía siempre una mezcla de células de dWL (verde) y células de dWP (rojo), independientemente del umbral utilizado. Esto sugería que las características elegidas no eran útiles para distinguir entre los 2 estadios a nivel celular (Fig. 6B). Sin embargo, la combinación de células de

dWL y dWP en el grafo principal presentaba una clara preferencia en la distribución de acuerdo al tipo de polígonos (Fig. 6C), a pesar de que el número de lados de las células no era una de las características consideradas para la correlación. De esto deducimos que el número de lados de una célula condiciona los valores de las 14 características analizadas dando una cierta restricción local a los tejidos. Este resultado nos impulsó a buscar diferencias a otro nivel; la distribución de polígonos que presentaban ambos tejidos.

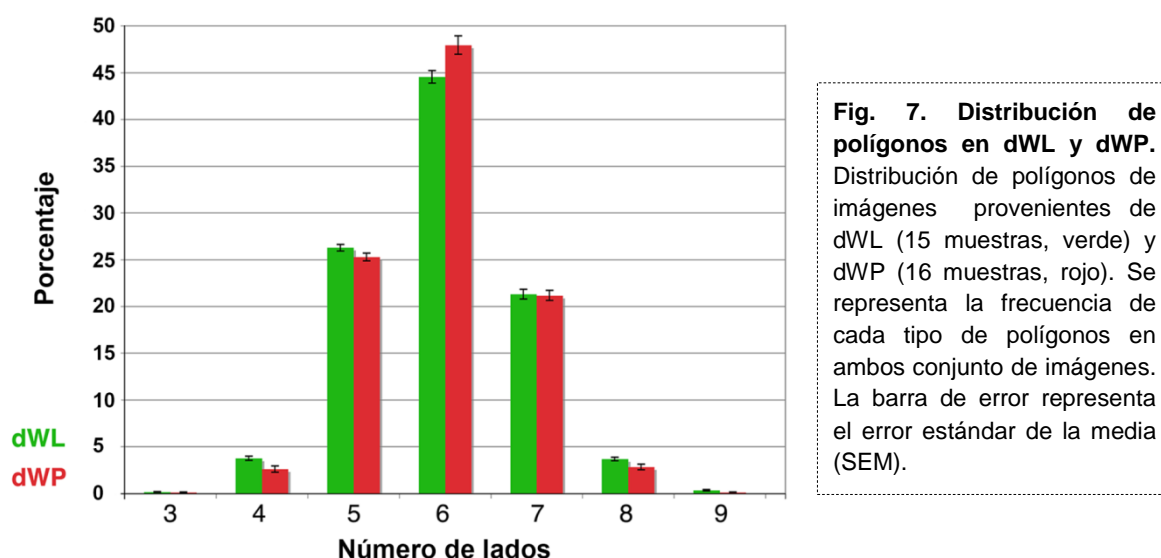
Categoría dWL (larva)		Categoría dWP (prepupa)	
Imagen	Número de células	Imagen	Número de células
1	354	1	637
2	465	2	691
3	412	3	480
4	439	4	326
5	420	5	369
6	816	6	558
7	958	7	636
8	505	8	608
9	858	9	416
10	792	10	341
11	516	11	303
12	713	12	422
13	666	13	333
14	633	14	360
15	523	15	206
		16	195
Total	9070	Total	6881

Tabla 2. Imágenes y células analizadas en las categorías dWL y en dWP. La tabla muestra el conjunto total de células extraídas en cada imagen analizada dentro de las categorías dWL y dWP. Se representa la suma total de células por categoría.

1.2. Estudio de la distribución de polígonos en el epitelio del ala en *Drosophila* durante el desarrollo larvario

Estudios previos han señalado que la aparición de una determinada distribución de polígonos en el disco de ala de *Drosophila* (alrededor 3% de células de 4 lados, 28% de pentágonos, 46% de hexágonos, 20% de heptágonos y 3% de octógonos) es una propiedad inherente de la proliferación en tejidos epiteliales y están presentes en otros metazoos como en la epidermis de la rana *Xenopus* o en la epidermis del cnidario de agua dulce *Hydra* (Gibson y cols., 2006). Esto está considerado como el primer nivel de organización en el plano del epitelio. Quisimos ver si nuestras imágenes cumplían con estas observaciones previas y por ello calculamos la distribución de polígonos en los 2 estadios de desarrollo analizados empleando 15 tejidos de dWL y 16 tejidos de dWP (Fig. 7). La distribución de polígonos para el total de imágenes se evaluó realizando la media de los datos y la barra de error se definió con el parámetro error estándar. En general, los valores medios de cada estadio fueron similares a los ya publicados

previamente (Gibson y cols., 2006; Aegerter-Wilmsen y cols., 2010). Sin embargo, observamos pequeñas diferencias en las distribuciones de dWL y dWP y quisimos cuantificarlo de forma precisa. Para ello, empleamos el test de análisis multivariado de varianza (MANOVA, del inglés “Multivariate Analysis of Variance”, Materiales y Métodos 3.2). Obtuvimos un p value de 0.013 que, al ser menor que el nivel de significancia 0.05, indica que se elimine la hipótesis nula de distribuciones significativamente iguales. En otras palabras, las distribuciones de polígonos presentaban diferencias significativas. Por ello, concluimos que durante el desarrollo de el disco de ala de *Drosophila* hay una pequeña variación en la distribución de polígonos durante el periodo de proliferación.



1.3. Estudio de características de grafos como medida de la organización del epitelio durante el desarrollo larvario

Nos planteamos si el cambio en la distribución de polígonos podría estar dirigiendo cambios progresivos en la organización epitelial y si estos cambios podrían ser detectados empleando diferentes metodologías. En un trabajo reciente, nuestro grupo había encontrado que diferencias en la organización epitelial pueden ser capturadas usando conceptos de teoría de grafos (Escudero y cols., 2011). Para aplicar estos métodos, diseñamos un set de 40 características (Tabla 1). Las primeras 8 características hacen referencia a propiedades geométricas de las células y las 32 restantes expresan conceptos de teoría de grafos y redes complejas (Costa y cols., 2007). Un análisis más profundo del significado de estas características lo podemos encontrar en Materiales y Métodos 2. Después de esto, empleamos el análisis de componentes principales (PCA, del inglés Principal Component Analysis, Materiales y

Métodos 3.3) para analizar los resultados. PCA es un algoritmo matemático no sesgado que reduce la dimensionalidad de los datos, tal que proyecta los datos sobre ejes ortogonales maximizando su dispersión. Por ello, las imágenes pueden ser representadas en un espacio bidimensional utilizando los 2 primeros componentes principales. En nuestro caso, los datos estaban formados por los valores de un subconjunto de características seleccionadas entre las 40 que definen las imágenes bajo estudio. Para elegir este subconjunto, diseñamos un paso de selección de características que extraiga solo aquellas más relevantes para discriminar entre los dos tipos de tejidos (Sáez y cols., 2013), (Materiales y Métodos 3.3). Un análisis de la gráfica resultante del PCA puede ser utilizado para identificar similitudes y diferencias entre imágenes. En este sentido, imágenes muy diferentes dentro de una misma categoría se distribuirán lejos unas de otras, mientras que imágenes muy similares se agruparán. Esta gráfica nos permitió cuantificar diferencias entre los grupos (dWL y dWP). Para estimar el grado de separación de manera cuantitativa usamos el “descriptor de PCA” (Materiales y Métodos 3.3) que nos dio un valor numérico para cada gráfico de PCA resultante de las diferentes combinaciones de características (en nuestro caso desde 0.1 a 1.8 aproximadamente). En el caso de la comparación dWL con dWP, en el paso de selección de características el programa eligió una característica geométrica y 4 de red: “Media Eje Mayor”, “Media Relación de Envolvente convexa de vecinos”, “D.E. Fuerza”, “Media Intermediación” y “Eficiencia”.

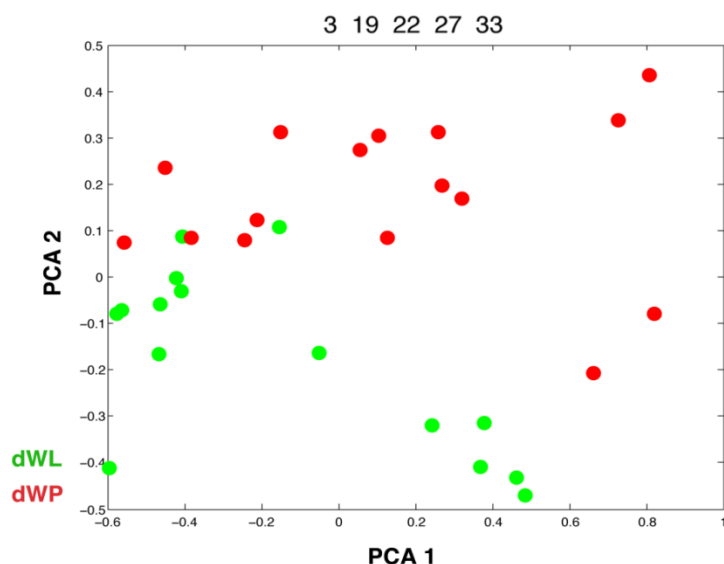


Fig. 8. PCA de imágenes dWL y dWP. Gráfico PCA para las comparaciones de imágenes dWL (puntos verdes) y dWP (puntos rojos) usando las características seleccionadas (3, 19, 22, 27 y 33).

Usando este subconjunto de características la gráfica de PCA mostró una separación entre los grupos dWL y dWP, con solo una pequeña región de solapamiento (Fig. 8). El valor correspondiente para el descriptor de PCA fue de 1.527. Los valores extraídos para estas características seleccionadas se muestran en Tabla 3.

dWL – Características seleccionadas

Imágenes	Media Eje Mayor	Media Relación de Envoltente convexa de vecinos	D.E. Fuerza	Media Intermediación	Eficiencia
1	63.85	0.964	84.8	5095.6	0.0030
2	55.73	0.970	78.0	7213.0	0.0031
3	55.83	0.970	81.6	6097.0	0.0032
4	54.80	0.971	74.2	6746.5	0.0033
5	61.52	0.968	83.3	6594.1	0.0030
6	41.38	0.978	60.4	15273.0	0.0031
7	44.92	0.975	63.1	14503.3	0.0031
8	46.14	0.978	62.1	11948.5	0.0030
9	42.52	0.971	59.4	16584.0	0.0031
10	42.23	0.977	56.6	14661.5	0.0031
11	50.71	0.977	61.5	7736.2	0.0031
12	43.69	0.979	62.2	13266.1	0.0031
13	46.24	0.975	57.6	12036.1	0.0030
14	47.84	0.975	60.5	11260.8	0.0030
15	53.15	0.972	67.5	8443.0	0.0030
Media	50.04	0.973	67.5	10497.2	0.0031
SEM	1.84	0.001	2.6	991.5	1.99881E-05

dWP – Características seleccionadas

Imágenes	Media Eje Mayor	Media Relación de Envoltente convexa de vecinos	D.E. Fuerza	Media Intermediación	Eficiencia
1	45.18	0.978	62.3	11038.1	0.0030
2	44.92	0.977	55.9	12043.5	0.0030
3	53.13	0.980	65.0	7437.4	0.0030
4	74.98	0.970	82.0	5075.9	0.0031
5	60.31	0.975	67.6	5572.4	0.0030
6	53.72	0.976	65.0	9346.4	0.0030
7	48.22	0.979	67.6	11316.4	0.0031
8	48.66	0.979	56.7	10392.1	0.0030
9	57.33	0.980	71.4	6314.3	0.0031
10	61.52	0.981	71.2	5202.0	0.0031
11	62.51	0.976	96.9	4160.5	0.0033
12	55.01	0.980	70.8	6391.7	0.0031
13	65.40	0.977	71.9	4780.2	0.0031
14	63.53	0.977	74.0	5461.1	0.0031
15	75.42	0.980	83.8	2728.2	0.0032
16	81.53	0.979	90.4	2353.8	0.0030
Media	59.46	0.978	72.0	6850.9	0.0031
SEM	2.74	0.001	2.8	769.8	1.88162E-05

Tabla 3. Valores de las características seleccionadas. Se muestran los valores que toman cada una de las características seleccionadas para cada imagen perteneciente a cada categoría (dWL y dWP). Se detalla el valor medio y el error estándar de la media (SEM) que toma cada característica seleccionada dentro de cada categoría. D.E. (Desviación Estándar).

Este resultado sugirió que nuestro método era capaz de identificar cambios topológicos entre dos estadios de desarrollo cercano. Sin embargo, existía la posibilidad que nuestro paso de selección de características pudiera estar forzando la separación entre ambos conjuntos de imágenes (restando relevancia biológica al experimento). Para descartar esta posibilidad, diseñamos un análisis para generar combinaciones aleatorias de nuestras imágenes y luego ejecutar el paso de selección de características y calcular el descriptor de PCA (Materiales y Métodos 3.4).

Los resultados para 4000 combinaciones mostraron que los valores obtenidos en la comparación de dWL con dWP era más alta que el 99.9% de las combinaciones aleatorias (Fig. 9). Solo 4 combinaciones aleatorias presentan un descriptor mayor que la comparación de dWL con dWP, representado en amarillo. Este último resultado apoya la existencia de diferencias reales y detectables en la organización de dWL y dWP.

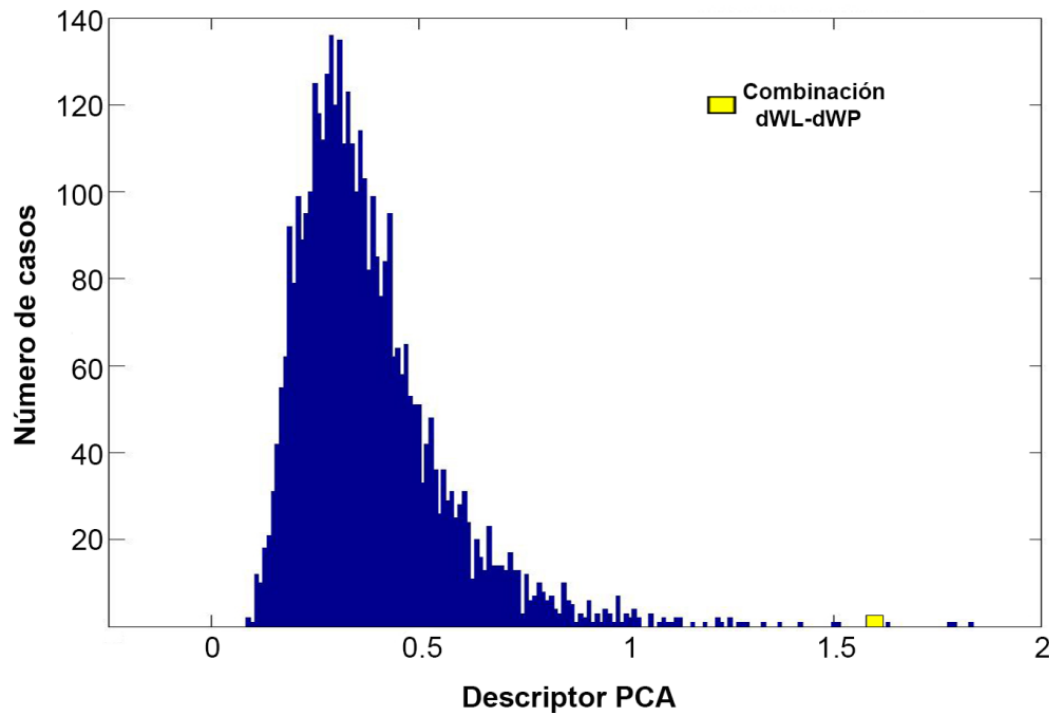


Fig. 9. Descriptor de PCA para 4000 combinaciones aleatorias de imágenes en 2 grupos. La gráfica muestra el número de combinaciones con un determinado valor del descriptor de PCA. Únicamente 4 combinaciones presentan un valor del descriptor de PCA mayor que nuestra combinación dWL-dWP (Cuadrado amarillo).

1.4. Relación entre los cambios en la distribución de polígonos y la aparición de una organización epitelial diferente.

Hasta aquí, hemos mostrado que las imágenes dWL y dWP presentan una distribución de polígonos y una organización epitelial diferentes. Una posibilidad sería que esto último fuera simplemente una consecuencia de una variación significativa en la distribución de polígonos. Para comprobarlo, realizamos un estudio de la organización del epitelio empleando el descriptor de PCA que nos indica el grado de separación atendiendo a la organización y un estudio de la distribución de polígonos empleando el test MANOVA que nos muestra si las distribuciones son significativamente diferentes para las mismas 4000 combinaciones de imágenes aleatorias.

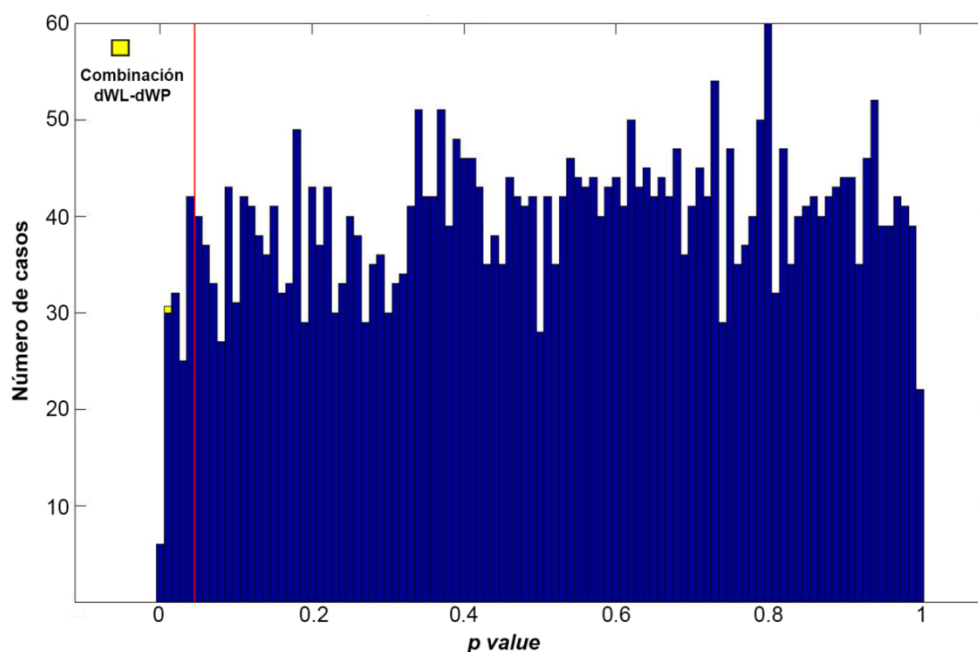


Fig. 10. Randomización de combinaciones de imágenes y p value del test MANOVA para sus respectivas distribuciones de polígonos. El gráfico muestra el número de combinaciones con una determinado p value. Se empleó el Test MANOVA para comparar las distribuciones de polígonos de cada combinación. Se representa el número de casos que presenta un determinado p value. El cuadrado amarillo se corresponde al p value de la combinación dWL-dWP.

El análisis de la distribución de polígonos mediante el test MANOVA nos mostró que solo en 32 casos (0.08%) el p value alcanzado era más pequeño que el obtenido en la comparación dWL con dWP (Fig. 10). En otras palabras, las distribuciones de polígonos de estos 32 casos presentaban una mayor diferencia que la comparación dWL con dWP. Posteriormente, representamos los p value del test MANOVA y los resultados del descriptor de PCA obtenidos en el apartado 1.3 (Fig. 11). La gráfica muestra que no existe una correlación entre la organización y la distribución de polígonos: Encontramos que los 4 casos con mayor diferencia en la organización (mejor descriptor PCA) no presentaban diferencias en la distribución de polígonos. Por otro lado, el valor del descriptor de PCA para los grupos arbitrarios con diferentes distribuciones de polígonos (MANOVA, p value < 0.05) eran menores que 1. Esto nos sugirió que, ambas características eran independientes entre sí.

Consecuentemente, solo en el caso de la comparación biológicamente significativa de dWL con dWP una gran diferencia en la organización (reflejado en un alto valor de descriptor de PCA) se correspondió con una diferencia significativa en la distribución de polígonos (reflejada en el bajo valor del p value del test MANOVA (Fig. 11)). Por ello pensamos en la existencia de un mecanismo de desarrollo que relaciona cambios en la distribución de polígonos con transformaciones detectables en la organización.

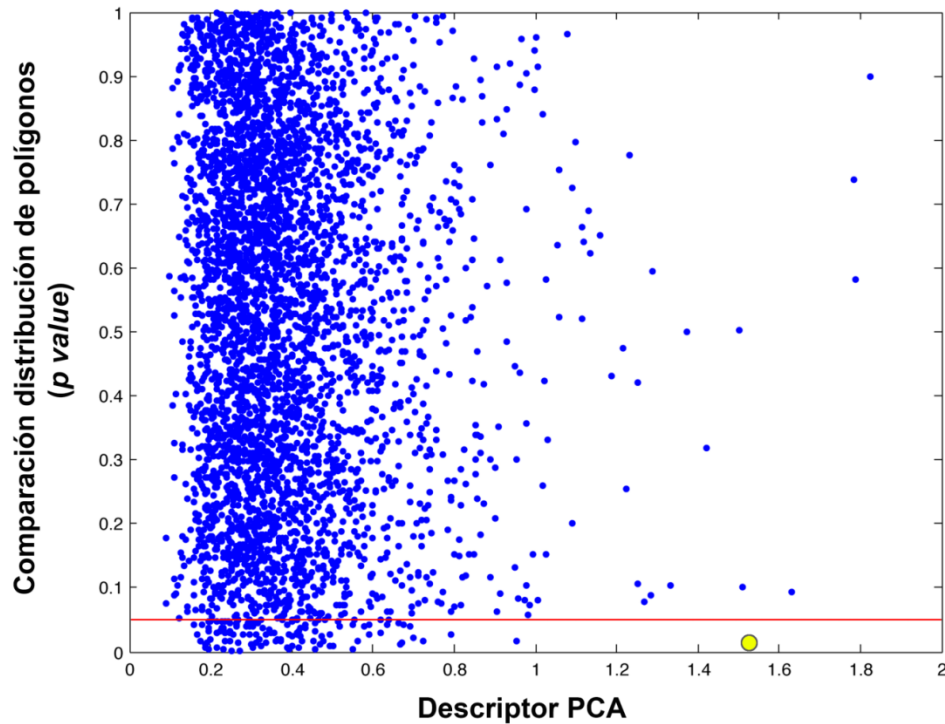


Fig. 11. Representación del descriptor PCA frente al p value del test MANOVA para distintas combinaciones de imágenes. El gráfico representa 4000 combinaciones aleatorias de imágenes (puntos azules). El p value, resultante de aplicar el test MANOVA para comparar las distribuciones de polígonos, se representa frente al descriptor PCA de las mismas combinaciones aleatorias. La línea roja marca p value=0.05. La gráfica muestra ausencia de correlación entre ambas variables. El círculo amarillo marca la combinación dWL-dWP.

2. Restricciones físicas de las células como mecanismo regulador de la organización de los tejidos empaquetados

En el primer capítulo se han evaluado los parámetros que influyen en la organización en tejidos epiteliales proliferativos usando el disco de ala de *Drosophila* como modelo. En este apartado ampliaremos nuestro estudio sobre la organización tisular a cualquier clase de tejido empaquetado presente en la naturaleza.

2.1. Análisis de la distribución de polígonos en diversos tejidos empaquetados

Como hemos mencionado en el capítulo anterior, se ha descrito que el epitelio en proliferación en metazoos presenta una "distribución de polígonos estereotipada" que surge, al menos en parte, de los mecanismos de división celular o de las tasas de crecimiento de estos tejidos (Gibson y cols., 2006; Aegerter-Wilmsen y cols., 2010; Nagpal y cols., 2008; Patel y cols., 2009). También hemos confirmado este hecho con nuestras propias imágenes de proliferación del ala de disco de *Drosophila* en estadio de prepupa (dWP, Sanchez-Gutierrez y cols., 2013), (Fig. 12A,E y Tabla 4). Con el objetivo de aclarar los mecanismos detrás de esta distribución estereotipada, se analizó la organización de un tejido no proliferativo, que también presenta células empaquetadas: las fibras musculares (Wigmore y Duglison, 1998). Para ello analizamos una serie de imágenes de bíceps control humano (BCA, Materiales y Métodos 1.2, Fig. 12B,C). Sorprendentemente, nos encontramos con la misma distribución de polígonos estereotipada (Fig. 12E y Tabla 4). También investigamos otros epitelios proliferativos de diferente procedencia. El análisis del neuroepitelio de pollo (cNT, Materiales y Métodos 1.1), (Frade, 2002; Wilcock y cols., 2007) presentaba una disposición celular diferente bajo una simple inspección visual. Las diferentes células que lo componen poseían superficies apicales pequeñas y grandes que se mezclaban dando como resultado un tejido muy heterogéneo (Fig. 12D). Curiosamente, la distribución de polígonos en cNT resultó muy diferente a la distribución estereotipada mencionada previamente, con un claro decremento en el número de hexágonos y un incremento de células con 4 y 8 lados (Fig. 12E). Estos resultados desafiaban el concepto de que todos los tejidos proliferativos presentan la distribución estereotipada.

2.2. Análisis de tejidos mediante el uso de conceptos geométricos

Los tejidos naturales empaquetados están caracterizados por estar formados por células organizadas como polígonos que rellenan la totalidad del tejido. El empleo de

conceptos geométricos en el análisis de tejidos puede ayudar a comprender los principios de arquitectura y modelado durante el desarrollo. Hemos usado la teselación de Voronoi (Voronoi, 1908) para crear diagramas que simulen tejidos naturales para mejorar la comprensión de la emergencia de organización en tejidos empaquetados. Estudios previos han demostrado que la existencia de una distribución al azar de semillas en el plano euclidiano produce un diagrama de Poisson-Voronoi con una

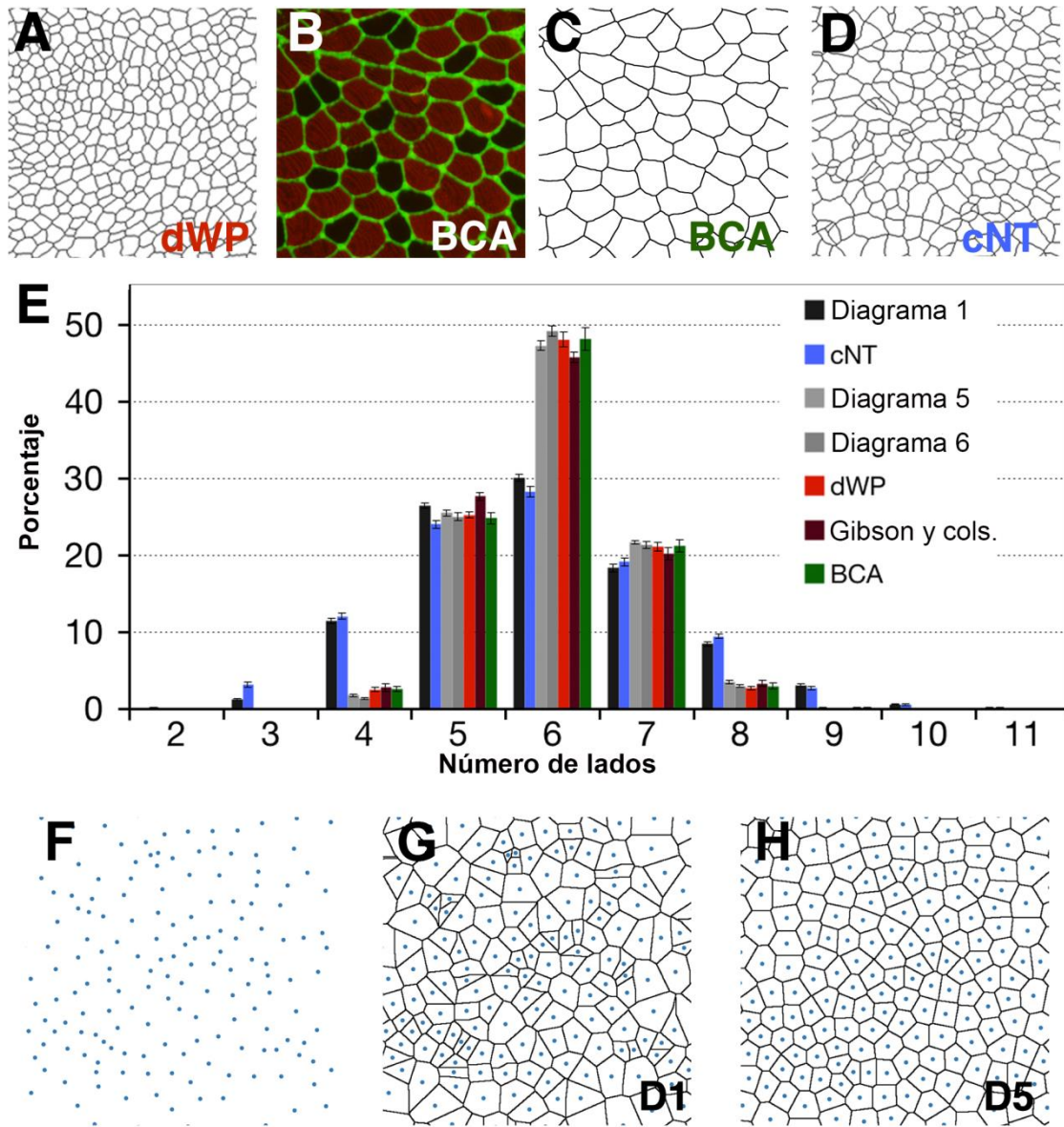


Fig. 12. Distribución de polígonos de tejidos biológicos comparados con diagramas de Voronoi. **A)** Segmentación de una imagen del disco imaginal del epitelio del ala de *Drosophila* en estado de prepupa. **B)** Biopsia de un bíceps adulto tratada con anticuerpos anti-colágeno VI (verde) y cadena pesada de la anti-miosina (rojo). **C)** Segmentación de la imagen mostrada en (B). **D)** Imagen segmentada del neuroepitelio de pollo. **E)** Distribución de polígonos de los distintos tejidos biológicos analizados. Se pueden distinguir 2 tipos de distribuciones: La distribución de polígonos de Poisson Voronoi para D1 y cNT, y la distribución de polígonos estereotipada para el resto de categorías. **F)** Diagrama con semillas distribuidas aleatoriamente en el plano. **G)** Teselación de Poisson-Voronoi formado a partir de las semillas indicadas en (A). **H)** Diagrama de Voronoi resultante después de aplicar 4 veces el algoritmo de Lloyd al diagrama 1. En todas las figuras los datos están representados con la media \pm el error estándar de la media (SEM).

Distribución de Polígonos

Tejidos	2	3	4	5	6	7	8	9	10	11	12	
Diagrama 1	0.00	1.11	11.13	26.11	30.40	18.63	8.66	3.15	0.64	0.14	0.03	Media
	0.00	0.13	0.36	0.37	0.47	0.49	0.30	0.20	0.09	0.05	0.02	Error
Diagrama 2	0.00	0.09	5.79	28.00	37.01	21.22	6.55	1.20	0.14	0.00	0.00	Media
	0.00	0.04	0.21	0.41	0.67	0.46	0.27	0.18	0.05	0.00	0.00	Error
Diagrama 3	0.00	0.00	3.69	27.15	41.92	21.53	5.11	0.55	0.05	0.00	0.00	Media
	0.00	0.00	0.19	0.48	0.55	0.55	0.26	0.08	0.03	0.00	0.00	Error
Diagrama 4	0.00	0.00	2.66	26.32	44.59	21.78	4.32	0.33	0.00	0.00	0.00	Media
	0.00	0.00	0.17	0.46	0.56	0.37	0.26	0.07	0.00	0.00	0.00	Error
Diagrama 5	0.00	0.00	1.77	25.56	47.29	21.70	3.53	0.16	0.00	0.00	0.00	Media
	0.00	0.00	0.17	0.40	0.63	0.23	0.19	0.04	0.00	0.00	0.00	Error
Diagrama 6	0.00	0.00	1.38	25.01	49.16	21.37	3.00	0.08	0.00	0.00	0.00	Media
	0.00	0.00	0.13	0.50	0.69	0.41	0.20	0.03	0.00	0.00	0.00	Error
Diagrama 10	0.00	0.00	0.56	22.65	54.45	20.86	1.45	0.03	0.00	0.00	0.00	Media
	0.00	0.00	0.09	0.25	0.51	0.42	0.13	0.02	0.00	0.00	0.00	Error
Diagrama 12	0.00	0.00	0.39	21.58	56.76	20.21	1.05	0.00	0.00	0.00	0.00	Media
	0.00	0.00	0.09	0.30	0.50	0.31	0.15	0.00	0.00	0.00	0.00	Error
Diagrama 20	0.00	0.00	0.14	18.84	62.43	18.06	0.53	0.00	0.00	0.00	0.00	Media
	0.00	0.00	0.04	0.35	0.64	0.35	0.10	0.00	0.00	0.00	0.00	Error
Diagrama 30	0.00	0.00	0.05	16.96	66.19	16.55	0.25	0.00	0.00	0.00	0.00	Media
	0.00	0.00	0.03	0.35	0.66	0.30	0.06	0.00	0.00	0.00	0.00	Error
Diagrama 40	0.00	0.00	0.03	16.01	67.89	15.86	0.21	0.00	0.00	0.00	0.00	Media
	0.00	0.00	0.02	0.32	0.58	0.31	0.05	0.00	0.00	0.00	0.00	Error
Diagrama 50	0.00	0.00	0.03	15.58	68.72	15.49	0.17	0.00	0.00	0.00	0.00	Media
	0.00	0.00	0.02	0.35	0.59	0.27	0.05	0.00	0.00	0.00	0.00	Error
Diagrama 100	0.00	0.00	0.00	15.59	68.99	15.30	0.12	0.00	0.00	0.00	0.00	Media
	0.00	0.00	0.00	0.32	0.65	0.44	0.04	0.00	0.00	0.00	0.00	Error
Diagrama 200	0.00	0.00	0.00	15.59	68.99	15.30	0.12	0.00	0.00	0.00	0.00	Media
	0.00	0.00	0.00	0.32	0.65	0.44	0.04	0.00	0.00	0.00	0.00	Error
dWL	0.00	0.12	3.73	26.19	44.37	21.61	3.66	0.32	0.01	0.00	0.00	Media
	0.00	0.04	0.23	0.37	0.57	0.34	0.18	0.06	0.01	0.00	0.00	Error
dWP	0.00	0.08	2.54	25.27	48.10	21.16	2.73	0.11	0.02	0.00	0.00	Media
	0.00	0.04	0.31	0.41	0.96	0.54	0.28	0.03	0.01	0.00	0.00	Error
dMWP	0.13	0.91	7.90	28.85	35.77	17.85	6.73	1.48	0.39	0.00	0.00	Media
	0.13	0.29	0.86	2.18	2.18	1.13	0.75	0.44	0.17	0.00	0.00	Error
cNT	0.18	3.21	12.09	24.04	28.30	19.18	9.48	2.74	0.60	0.13	0.04	Media
	0.07	0.32	0.41	0.46	0.74	0.51	0.32	0.23	0.10	0.05	0.02	Error
EYE	0.00	0.58	14.43	24.73	28.12	19.17	9.07	3.11	0.67	0.12	0.00	Media
	0.00	0.13	0.61	1.54	1.60	1.00	0.57	0.26	0.22	0.12	0.00	Error
Volvox	0.00	0.00	0.50	21.50	57.00	19.00	1.50	0.00	0.00	0.00	0.00	Media
	-	-	-	-	-	-	-	-	-	-	-	Error
BCA	0.00	0.00	2.61	24.84	48.18	21.22	3.00	0.16	0.00	0.00	0.00	Media
	0.00	0.00	0.31	0.73	1.52	0.77	0.39	0.09	0.00	0.00	0.00	Error
BNA	0.00	0.24	4.96	26.14	36.92	24.62	5.77	1.36	0.00	0.00	0.00	Media
	0.00	0.16	1.02	1.76	1.95	1.70	0.98	0.50	0.00	0.00	0.00	Error
Simulación Control sin proliferación	0.00	0.00	1.22	25.44	50.62	20.20	2.48	0.03	0.00	0.00	0.00	Media
	0.00	0.00	0.22	0.51	0.86	0.47	0.27	0.03	0.00	0.00	0.00	Error
Simulación Atrofia	0.03	0.05	1.55	25.67	48.57	21.59	2.35	0.16	0.03	0.00	0.00	Media
	0.03	0.05	0.22	0.53	0.82	0.49	0.29	0.07	0.03	0.00	0.00	Error
Simulación Área Ideal	0.00	0.00	1.27	24.42	51.06	20.89	2.23	0.12	0.00	0.00	0.00	Media
	0.00	0.00	0.19	0.65	0.97	0.58	0.27	0.06	0.00	0.00	0.00	Error
Simulación Control con proliferación	0.00	0.00	0.55	25.67	50.26	21.36	2.13	0.04	0.00	0.00	0.00	Media
	0.00	0.00	0.09	0.33	0.58	0.40	0.17	0.02	0.00	0.00	0.00	Error
Caso 2	0.00	0.00	0.34	24.58	52.85	20.41	1.71	0.10	0.00	0.00	0.00	Media
	0.00	0.00	0.09	0.31	0.60	0.43	0.19	0.05	0.00	0.00	0.00	Error
Caso 3	0.00	0.00	5.67	34.74	29.62	19.94	8.79	1.24	0.00	0.00	0.00	Media
	0.00	0.00	0.50	0.66	0.97	0.48	0.34	0.21	0.00	0.00	0.00	Error
Caso 4	0.00	0.00	3.27	36.07	35.29	14.95	8.26	2.10	0.07	0.00	0.00	Media
	0.00	0.00	0.39	0.68	0.97	0.55	0.38	0.26	0.05	0.00	0.00	Error

Tabla 4. Distribución de polígonos de tejidos naturales y de patrones geométricos. Valores medios de cada tipo de polígono en todos los tipos de teselaciones analizados. También se representa su error estándar. El caso de volvox no dispone de error ya que se utilizaron los datos publicados de una sola muestra (Korn y Spalding, 1973).

distribución poligonal fija que es independiente de la cantidad de semillas de partida (Miles y Maillardet, 1982). Hemos generado veinte de estos diagramas en laboratorio y reproducido esos resultados (Fig. 12F,G) obteniendo histogramas con hexágonos (29,5%), pentágonos (25,9%) y heptágonos (19,9%), junto con otros polígonos menos frecuentes (Fig. 12E Diagrama 1 y Tabla 4). Un diagrama de Voronoi puede ser homogeneizado mediante la aplicación del algoritmo de Lloyd (Du y cols., 1999; Lloyd, 1982). Este proceso identifica los centroides de las células de Voronoi y los usa como semilla para llevar a cabo un nuevo diagrama de Voronoi produciendo una relajación de este diagrama. Esto hace que las células sean más homogéneas en términos de forma y tamaño celular tras cada iteración (Fig. 12H). La aplicación iterativa del algoritmo de Lloyd aproxima la imagen a un diagrama de Voronoi óptimo (Du y cols., 1999; Lloyd, 1982), donde la semilla de cada célula de Voronoi es su centroide (CVT, del inglés "Centroidal Voronoi Tessellation", Materiales y Métodos 4.1). A partir de los diagramas de Poisson-Voronoi generados, obtuvimos la distribución de polígonos de 200 iteraciones de Lloyd y analizamos su distribución de polígonos (Fig. 13A y Tabla 4). Los histogramas cambiaron rápidamente después de las primeras iteraciones, aumentando el número de hexágonos y heptágonos. Después de 50 iteraciones los histogramas se estabilizaron mostrando 70% de hexágonos. La combinación de estos diagramas y distribuciones forman lo que hemos llamado "vía CVT". Observamos que los diagrama 5 y 6 de la vía CVT coinciden con la distribución de polígonos estereotipada descrita previamente (Fig. 12E). Sin embargo, estos diagramas de Voronoi se obtuvieron sin necesidad de invocar ningún mecanismo de división celular. Nuestros análisis con las muestras de músculos y los diagramas de Voronoi sugirieron que la distribución de polígonos estereotipada no es una consecuencia topológica exclusiva de la mitosis y la proliferación celular.

A causa de la creación de los diagramas de Voronoi obtuvimos otros sorprendentes resultados. Primero, la distribución de polígonos de cNT era bastante similar a la distribución obtenida en el diagrama 1 de Voronoi (Fig. 12E y Tabla 4). Por otra parte, un examen detallado del activamente proliferativo del disco de ala en tercer instar (ver dWL en apartado 1.2 y Sanchez-Gutierrez y cols., 2013), nos mostró una distribución de polígonos más cercana a un diagrama 4 de Voronoi (Tabla 4). Otro ejemplo de tejido proliferante que no presenta la distribución de polígonos estereotipada, es *Volvox*, un alga verde que forma colonias esféricas que presentaban una proporción muy alta de hexágonos (Korn y Spalding, 1973), y una distribución similar a la presentada por el diagrama 12 de Voronoi (Tabla 4). Tomados en conjunto, estos resultados indicaron que los mecanismos de proliferación celular no podían explicar completamente la distribución de polígonos estereotipada. Más importante aún, habíamos encontrado un

fenómeno más general: los tejidos naturales analizados en esta y otras obras (Aegerter-Wilmsen y cols., 2010; Gibson y cols., 2006; Korn y Spalding, 1973; Lewis, 1928; Sanchez-Gutierrez y cols., 2013) siempre eran similares a una de las distribuciones producidos por alguno de los diagramas de la vía CVT (Fig. 12E). Sugiriendo que los tejidos naturales empaquetados no pueden presentar infinitas organizaciones, sino que están limitadas a ciertas combinaciones de distribuciones de polígonos.

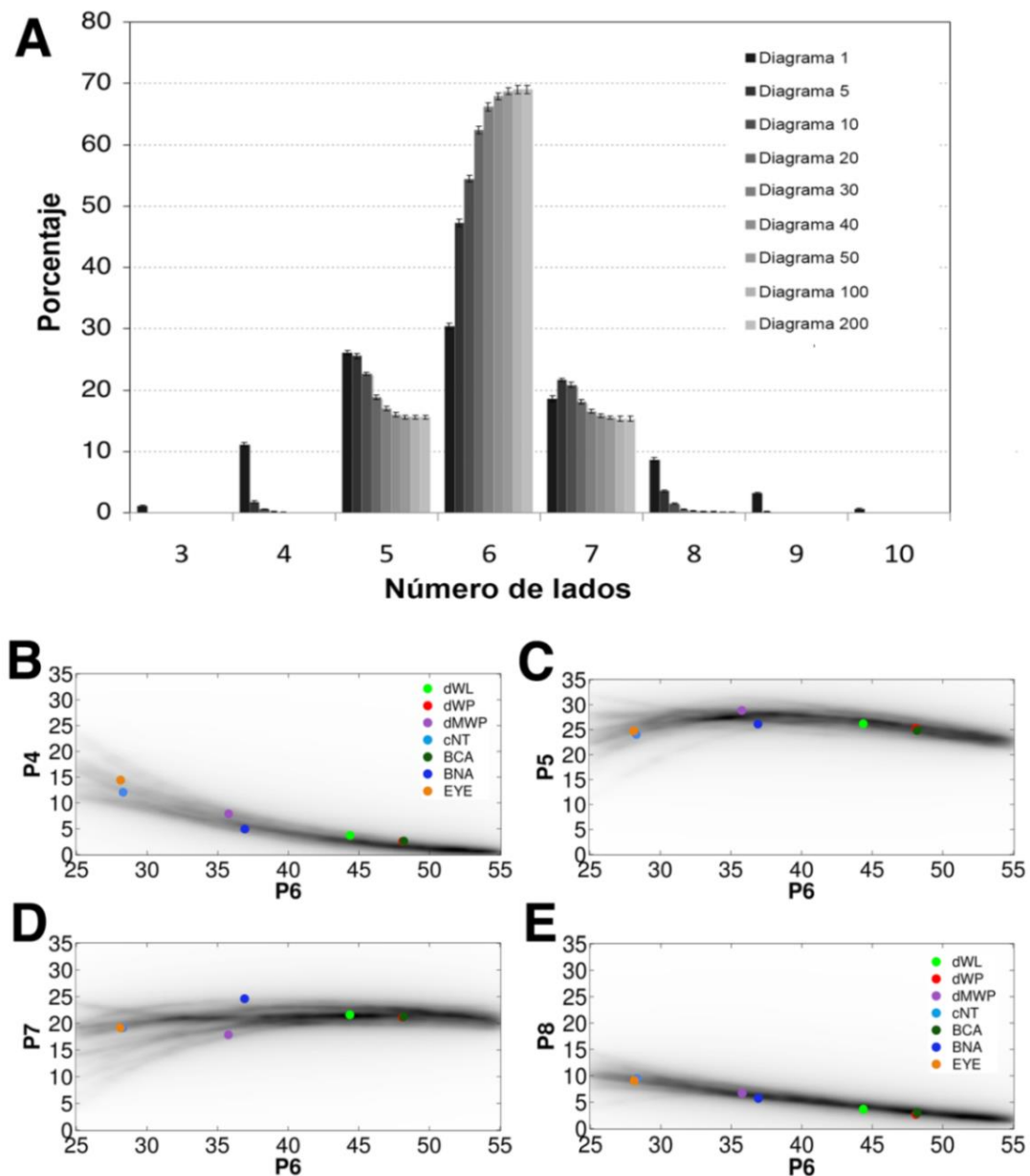


Fig. 13. Comparación de la distribución de polígonos de tejidos naturales con la vía CVT. A) Distribución de polígonos de los diagramas de Voronoi resultantes de la aplicación iterativa del algoritmo de Lloyd. **B-E)** Gráfica de densidad donde se muestra la frecuencia de células de cuatro, cinco, siete y ocho lados respectivamente, dependiente del porcentaje de hexágonos (P6) en los diagramas de CVT. La región más oscura refleja mayor probabilidad de ocurrencia de un valor determinado. Los círculos de colores representan la media de valores de cada clase de polígonos para los diferentes conjuntos de imágenes analizadas en este estudio.

Para comprobar esta hipótesis, diseñamos un método para cuantificar la similitud entre los tejidos provenientes de muestras naturales y la vía CVT. En cada uno de los diagramas de Voronoi el número de hexágonos tiene una correspondencia con el valor que toman los otros tipos de polígonos. Usamos esta propiedad para transformar los diferentes pasos de la vía CVT a una función continua (Materiales y Métodos 4.1.1). De este modo, pudimos determinar cómo varía el número de células con cuatro, cinco, siete u ocho lados según el porcentaje de hexágonos y convertir estos datos en una nube de probabilidad que facilitó la visualización de la similitud entre tejidos naturales y diagramas de Voronoi. (Fig. 13B-E). Esto nos permitió comparar cuantitativamente las diferentes imágenes naturales con la vía CVT y ver cómo se desvían de ésta. Confirmamos que imágenes provenientes de cNT, dWL, dWP y BCA presentaban una organización como la predicha por Voronoi utilizando los test estadísticos adecuados que nos permitieron evaluar diferencias entre las muestras (con pequeñas variaciones en polígonos menos frecuentes, Materiales y Métodos 4.1.2, Tabla 5). De ahí que, consideramos que la vía CVT servía como una escala objetiva para comparar la organización de cualquier tejido empaquetado a nivel de distribución de polígonos.

<i>p value</i>	P6-P4	P6-P5	P6-P7	P6-P8
dWL	0.002*	0.429	0.780	0.011*
dWP	0.012*	0.948	0.525	0.025*
dMWP	0.249	0.615	0.016	0.468
cNT	0.294	0.156	0.320	0.810
BCA	0.009*	0.483	0.702	0.376
BNA	0.308	0.115	0.031	0.495
EYE	0.437	0.829	0.556	0.716

Tabla 5. Comparación tejidos naturales con CVT continuo. Cada celda presenta el *p value* obtenido al emplear los tests de medias t-student, welch o wilcoxon (verde oscuro, verde medio y verde claro respectivamente) según la naturaleza de las muestras. El test de media evalúa la similitud en la proporción de cada tipo de polígonos que presenta un tejido con el que debería presentar (tomando como referencia los valores de la vía CVT) para el porcentaje de hexágonos suministrado por dicho tejido. El valor en negrita señala que la proporción del tipo de polígonos del tejido indicado es significativamente diferente de la vía CVT para ese tipo de polígonos. El valor con “*” se corresponde con porcentaje de polígonos poco frecuente y por tanto más variable.

2.3. Restricciones geométricas en los tejidos empaquetados

Tanto los diferentes tejidos naturales analizados como los diagramas de Voronoi compartían algunas similitudes: Las células eran similares a polígonos convexos que cubren el plano sin superponerse o dejar espacios vacíos. Este empaquetamiento de células poligonales ha sido estudiado desde distintos puntos de vista y es sabido que están sujetas a varias generalizaciones o leyes:

- ❖ El teorema de Euler: Dado un tejido con un gran número de células, el número medio de vecinos de cada célula debe ser seis (Weire y Rivier, 1984).
- ❖ La ley de Lewis afirma que la superficie media (A_n) de una célula con n -lados aumenta de forma lineal con n (así, las células pequeñas tienden a tener menos vecinos, mientras que las grandes tienden a tener más), (Chiu, 1995; Sangwoo y cols., 2014).
- ❖ Ley Aboav-Weaire afirma que las células con un mayor número de lados tienden a tener vecinos con pocos lados y viceversa (Chiu, 1995; Aboav, 1970).

Los tejidos analizados en este trabajo cumplían en gran medida con las 3 leyes descritas anteriormente (Fig. 14 y Tabla 6). Estas leyes implicaban una relación entre la distribución de área y la distribución de vecinos de las células. Es por ello que analizamos la distribución del área celular tanto en los distintos diagramas de Voronoi como en los tejidos naturales analizados (Fig. 15A,B, Materiales y Métodos 5).

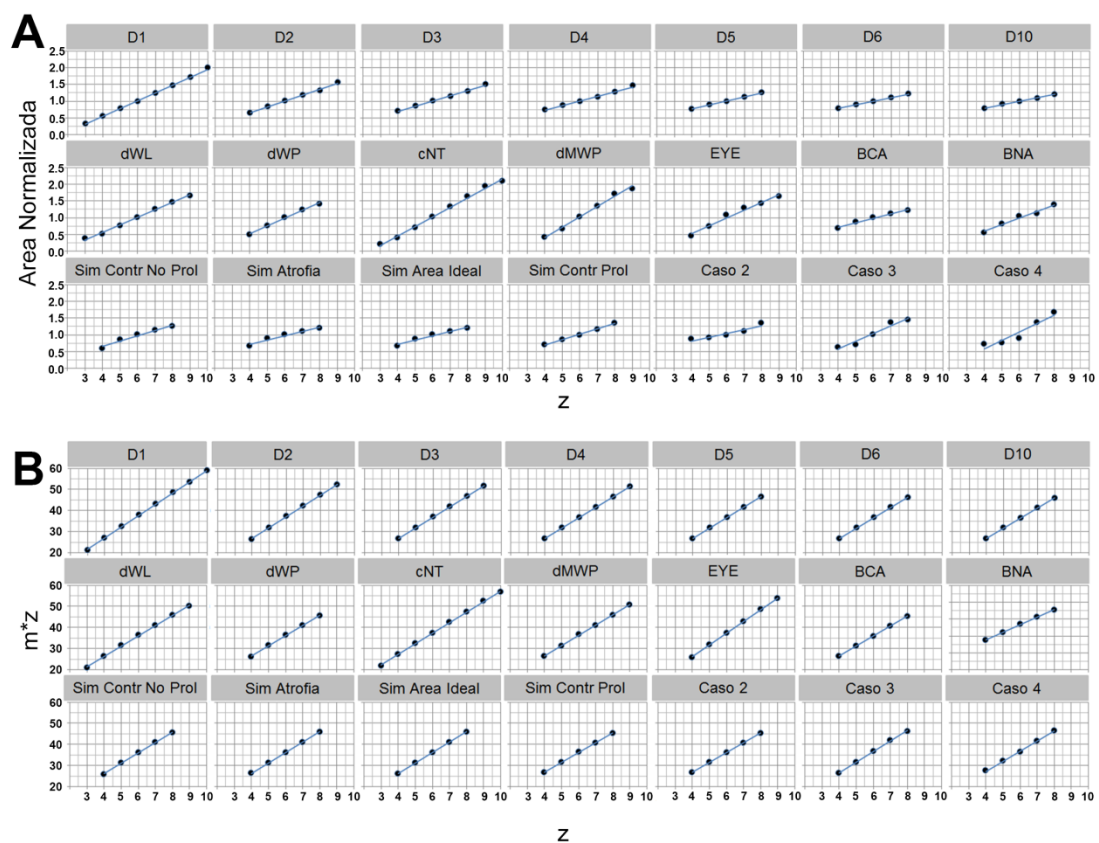


Fig. 14 Leyes geométricas que definen a los tejidos empaquetados y diagramas de Voronoi. A) Ley de Lewis. Todas las muestras indican una relación lineal entre el área de la célula y el número de lados (z). Cada punto representa la media del área celular a lo largo de todas las muestras. La línea azul indica el ajuste lineal de los datos. **B)** Ley de Aboav-Weaire. Todas las muestras presentan una relación lineal entre el número de lados de la célula (z) y el producto de esta cantidad con la media del número de lados de los vecinos de una célula (m).

Teorema de Euler

Tejidos	Media	D.E.	SEM
Diagrama 1	5.972	0.020	0.004
Diagrama 2	5.986	0.022	0.005
Diagrama 3	5.991	0.023	0.005
Diagrama 4	5.997	0.021	0.005
Diagrama 5	6.001	0.020	0.005
Diagrama 6	5.998	0.016	0.004
Diagrama 10	6.001	0.011	0.002
dWL	5.959	0.013	0.003
dWP	5.964	0.017	0.004
dMWP	5.894	0.060	0.020
cNT	5.911	0.016	0.004
EYE	5.946	0.004	0.003
BCA	5.976	0.028	0.005
BNA	6.035	0.071	0.021
Simulación Control sin proliferación	5.972	0.015	0.003
Simulación Atrofia	5.977	0.020	0.005
Simulación Área Ideal	6.055	0.020	0.004
Simulación Control con proliferación	5.990	0.006	0.001
Caso 2	5.989	0.007	0.002
Caso 3	5.952	0.032	0.008
Caso 4	5.954	0.023	0.006

Tabla 6. Teorema de Euler. Se muestra el número medio de vecinos por célula en cada una de las teselaciones analizadas en este estudio. Adicionalmente se representa la desviación estándar (D.E.) y el error estándar de la media (SEM).

En los diagramas de Voronoi, la distribución de área del diagrama 1 estaba sesgada a la izquierda como resultado de un alto número de células con área pequeña presente en estas imágenes. En los sucesivos diagramas, la distribución de área presentó un mayor grado de simetría con una base más estrecha, debido al incremento de homogeneidad entre las células (Fig. 15A). Los tejidos naturales proliferativos, dWL y dWP mostraban una distribución de área más simétrica debido al incremento de la homogeneidad en el tamaño de las células; mientras que cNT presentaba una distribución de área sesgada a la izquierda debido al alto número de células con pequeña superficie apical (Fig. 15B y Tabla 7). En el caso de tejidos de músculos no proliferativos, observamos que BCA presentaba la misma distribución de área que Diagrama 5 (Fig. 15B y Tabla 7). Esto estaría reflejando la ausencia de varianza en el tamaño de las células debido a la proliferación.

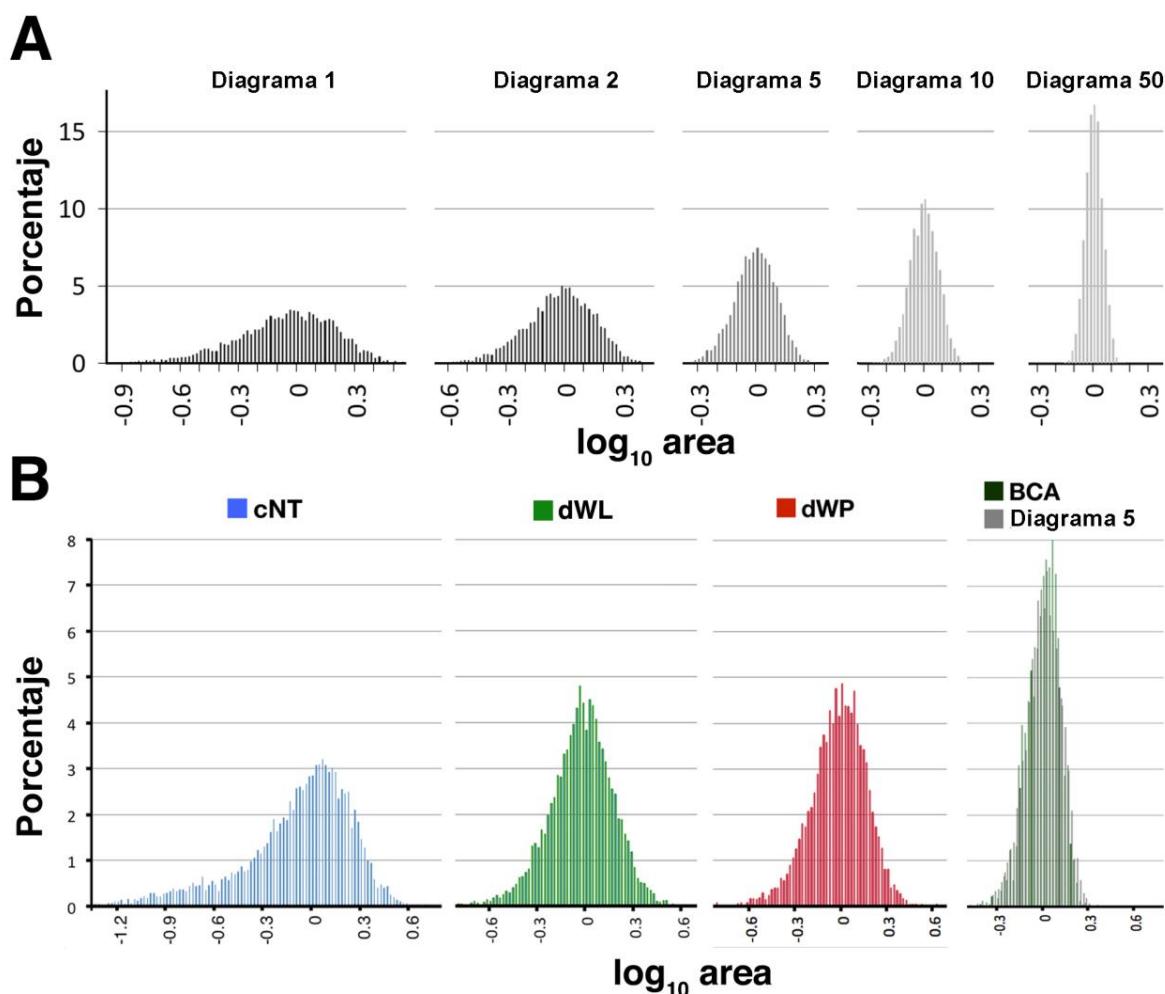


Fig. 15. Distribución de área de tejidos naturales y geométricos. A) Comparación de la distribución del área celular de varios diagramas de Voronoi a lo largo de la vía CVT. El área de las células se vuelve progresivamente más homogénea tras cada iteración de Lloyd. **B)** Comparación de la distribución del área celular para epitelios proliferativos cNT (azul), dWL (verde) y dWP (rojo). A la derecha tenemos representado Bíceps (verde oscuro) junto con diagrama 5 de Voronoi (gris).

Por tanto observamos que un aumento de la homogeneidad estaba relacionado con un aumento en el porcentaje de hexágonos, pentágonos y heptágonos. En los tejidos más heterogéneos como cNT el gran número de células pequeñas aumenta directamente el número de células con pocos lados (ley de Lewis), pero indirectamente también proporciona lados a las células más grandes (ley Aboav-Weaire). En resumen, nuestros resultados apoyaban la existencia de unas restricciones que afectan a cualquier tejido natural compuesto por polígonos convexos y que hace que correlacione la forma de la distribución de área de las células con la frecuencia con la que emergen los distintos tipos de polígonos.

Distribución de área - Estadísticos

Tejidos	Media	Varianza	D.E.	Moda	Mediana	Asimetría
Diagrama 1	-0.069	0.068	0.260	-0.207	-0.042	-0.698
Diagrama 2	-0.033	0.031	0.175	-0.249	-0.021	-0.442
Diagrama 3	-0.022	0.020	0.141	-0.055	-0.013	-0.346
Diagrama 4	-0.016	0.015	0.121	-0.180	-0.010	-0.281
Diagrama 5	-0.013	0.012	0.108	-0.122	-0.009	-0.234
Diagrama 10	-0.007	0.006	0.076	-0.014	-0.006	-0.105
Diagrama 50	-0.002	0.002	0.046	-0.021	-0.0023	-0.035
Diagrama 100	-0.002	0.002	0.046	0.008	-0.0023	-0.044
Diagrama 200	-0.002	0.002	0.046	0.008	-0.0023	-0.044
dWL	-0.044	0.041	0.202	-0.037	-0.036	-0.538
dWP	-0.037	0.035	0.186	-0.212	-0.023	-0.562
cNT	-0.106	0.117	0.342	-0.103	-0.035	-1.033
dMWP	-0.099	0.097	0.311	-0.470	-0.074	-0.800
EYE	-0.105	0.103	0.321	-0.722	-0.035	-0.394
BCA	-0.014	0.013	0.112	-0.209	-0.001	-0.736
BNA	-0.053	0.060	0.246	-0.105	0.003	-1.659
Sim. Control sin proliferación	-0.004	0.004	0.061	-0.099	0.004	-1.196
Sim. Atrofia	-0.010	0.014	0.117	0.025	0.016	-11.450
Sim. Área Ideal	-0.006	0.005	0.073	-0.106	-0.012	0.433
Sim. Control con proliferación	-0.007	0.006	0.075	-0.084	-0.018	2.044
Caso 2	-0.006	0.005	0.068	-0.090	-0.019	2.355
Caso 3	-0.034	0.030	0.172	-0.234	-0.139	0.133
Caso 4	-0.030	0.024	0.154	-0.179	-0.109	0.866

Distribución de área - Percentiles

Tejidos	Per.0	Per.1	Per.10	Per.25	Per.50	Per.75	Per.90	Per.99	Per.100
Diagrama 1	-1.482	-0.798	-0.407	-0.221	-0.042	0.119	0.235	0.417	0.607
Diagrama 2	-0.816	-0.496	-0.264	-0.141	-0.021	0.091	0.182	0.322	0.456
Diagrama 3	-0.582	-0.392	-0.207	-0.110	-0.013	0.077	0.154	0.267	0.389
Diagrama 4	-0.489	-0.323	-0.178	-0.094	-0.010	0.067	0.136	0.233	0.348
Diagrama 5	-0.438	-0.275	-0.156	-0.082	-0.009	0.063	0.122	0.212	0.314
Diagrama 10	-0.281	-0.188	-0.104	-0.059	-0.006	0.047	0.091	0.162	0.246
Diagrama 50	-0.159	-0.112	-0.061	-0.034	-0.0023	0.028	0.058	0.100	0.157
Diagrama 100	-0.159	-0.111	-0.061	-0.033	-0.0023	0.028	0.057	0.099	0.153
Diagrama 200	-0.159	-0.111	-0.061	-0.033	-0.0023	0.028	0.057	0.099	0.153
dWL	-1.900	-0.586	-0.299	-0.163	-0.036	0.087	0.200	0.389	0.912
dWP	-1.153	-0.560	-0.275	-0.147	-0.023	0.089	0.182	0.349	0.614
cNT	-1.523	-1.146	-0.599	-0.266	-0.035	0.131	0.257	0.458	0.771
dMWP	-2.443	-0.970	-0.484	-0.278	-0.074	0.115	0.267	0.531	0.795
EYE	-1.060	-0.848	-0.553	-0.385	-0.035	0.154	0.279	0.427	0.769
BCA	-0.680	-0.329	-0.156	-0.086	-0.001	0.067	0.119	0.195	0.238
BNA	-1.188	-0.879	-0.401	-0.114	0.003	0.090	0.188	0.304	0.345
Sim. Control sin proliferación	-0.502	-0.185	-0.081	-0.036	0.004	0.037	0.064	0.104	0.142
Sim. Atrofia	-2.790	-0.295	-0.138	-0.032	0.016	0.047	0.073	0.109	0.142
Sim. Área Ideal	-0.397	-0.162	-0.088	-0.048	-0.012	0.024	0.111	0.177	0.202
Sim. Control con proliferación	-0.220	-0.113	-0.078	-0.048	-0.018	0.010	0.050	0.282	0.349
Caso 2	-0.115	-0.089	-0.062	-0.044	-0.019	0.002	0.055	0.260	0.323
Caso 3	-0.299	-0.258	-0.222	-0.194	-0.139	0.150	0.170	0.192	0.199
Caso 4	-0.221	-0.191	-0.164	-0.139	-0.109	0.147	0.224	0.243	0.250

Tabla 7. Estadísticos y percentiles de la distribución de área. Los estadísticos se extraen de la distribución del área normalizada. La media, varianza, D.E., moda, mediana y asimetría son estadísticos que determinan numéricamente el comportamiento de la distribución. El percentil (Per) nos muestra el valor del área normalizada bajo el cual se encuentra un porcentaje dado de observaciones. D.E. (Desviación Estándar).

2.4. Restricciones biofísicas en los tejidos empaquetados

Nos preguntamos por qué la vía CVT era capaz de predecir los diferentes tipos de distribuciones de polígonos encontrados en los tejidos naturales. Una posible explicación es que los diagramas de Voronoi estuvieran reflejando algunas de las propiedades biofísicas celulares de las células epiteliales (Farhadifar y cols., 2007). En términos simplistas, las “células” en los diagramas de Voronoi emergen de semillas que crecen ‘empujando’ unas a otras con igual fuerza hasta encontrarse en una distancia media para formar el contorno de la “célula” de Voronoi. Una posibilidad es que las células compartieran esta homogeneidad de fuerzas tal y como dicta la forma que tienen de empaquetarse. Para comprobarlo, primero analizamos un tejido donde las fuerzas de adherencia y contractilidad eran claramente asimétricas: el disco imaginal de ojo de *Drosophila* (Escudero y cols., 2007), (Fig. 16A). Durante el desarrollo del ojo, los fotoreceptores forman agrupaciones con pequeñas áreas apicales rodeadas por 4 células como más grandes. En este tejido la frecuencia de los tipos de polígonos fue similar al Diagrama 1 de Voronoi (Fig. 17). Consecuentemente, la distribución de áreas celulares presentaba una alta varianza y estaba sesgada por la izquierda, aunque en la forma de una distribución bimodal de células pequeñas y grandes (Fig. 16B y Tabla 7). Aunque hay una clara organización asimétrica en este tejido, la distribución de polígonos permanecía en los límites de la vía CVT (Fig. 13B-E). Estos resultados sugirieron que los cambios fisiológicos que se producían de forma balanceada en la contractilidad y/o adhesión estaban incrementando los niveles de heterogeneidad en el área de las células, modificando pero manteniendo, el tejido dentro de los límites de la vía CVT.

Estudiamos entonces el efecto de alteraciones no fisiológicas de propiedades biofísicas de la célula. Para ello analizamos tanto alteraciones genéticas como patológicas. La miosina II es una molécula contráctil responsable de la arquitectura de las células epiteliales (Pilot y Lecuit, 2005). Su desregulación altera la contracción cortical y adherencia de la zona apical de las células epiteliales, así como el proceso de citocinesis (Escudero y cols., 2007; Young y cols., 1993). Analizamos las consecuencias de una leve reducción relativamente homogénea de la miosina II en el epitelio de ala de *Drosophila* (Materiales y Métodos 1.1, Escudero y cols., 2011). El epitelio resultante (dMWP) presentaba células con tamaños y formas heterogéneas basándonos en los contornos apicales (Fig. 16C). Esta heterogeneidad en la reducción de miosina II fue también la causa de la alta diversidad entre las diferentes muestras analizadas.

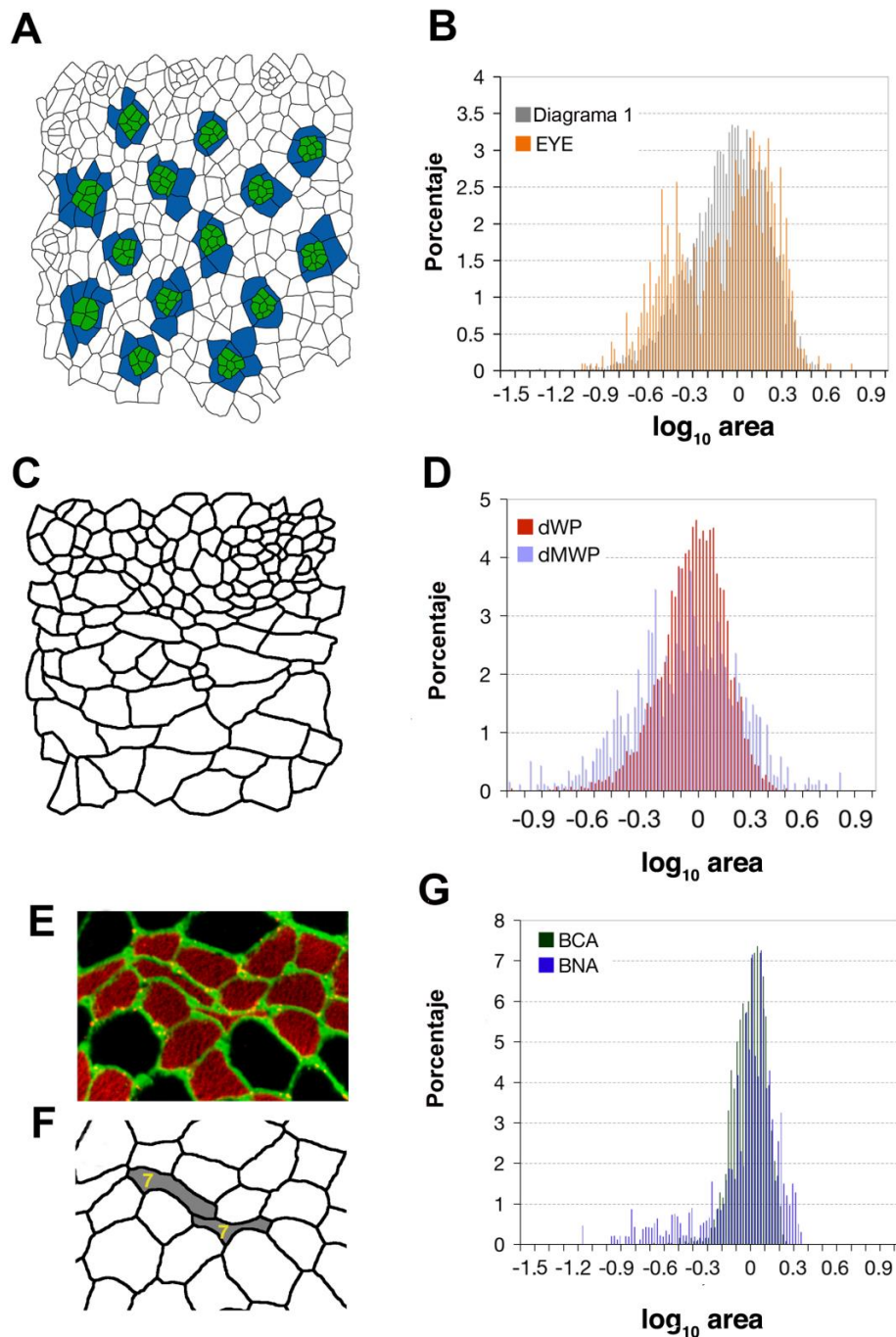


Fig. 16. Comparación distribución de área de tejidos alterados con tejidos naturales y diagrama de Voronoi. Tejido alterado: EYE. A) Segmentación de disco imaginal de ojo de *Drosophila* en estado de prepupa. Los fotoreceptores en desarrollo analizados están marcados en verde. Las células adyacentes a los fotoreceptores están marcadas en azul. **B)** Comparación de la distribución del diagrama 1 (gris) y EYE (naranja). EYE presentan una distribución bimodal de área sesgada a la izquierda. **Tejido alterado: dMWP. C)** Segmentación de disco imaginal de ala de *Drosophila* en estado de prepupa donde los niveles de miosina II han sido reducidos. **D)** Comparación de la distribución de área de dWP (rojo) y dMWP (violeta). El ala mutante presenta una frecuencia más irregular de tamaño y está ligeramente sesgada a la izquierda. **Tejido alterado: BNA. E)** Biopsia de bíceps de adulto proveniente de un paciente con la patología de atrofia neurógena. Las fibras están marcadas con el anticuerpo colágeno VI (verde). **F)** Segmentación de la imagen mostrada en (E). **G)** Comparación de la distribución de área de BCA (verde oscuro) y BNA (azul). El alto número de pequeñas fibras atróficas, que son características de la patología de atrofia neurógena, produce una distribución sesgada a la izquierda en el caso de BNA.

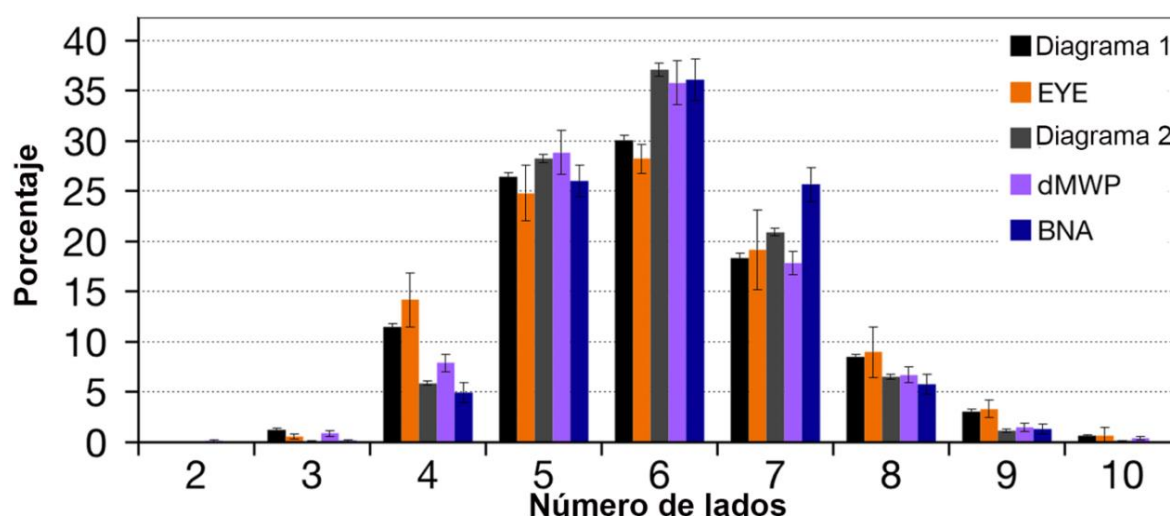


Fig. 17. Distribución de polígonos de tejidos alterados. Comparación de la distribución de polígonos de diferentes tejidos en los que la contractibilidad, la tensión o el volumen de estabilidad no son homogéneos en todo el tejido.

En comparación con el epitelio de ala de *Drosophila* control (dWP), el epitelio dMWP tenía una distribución de áreas celulares más irregular, sesgada moderadamente a la izquierda y con una varianza mayor (Fig. 16D y Tabla 7). Como resultado, se alteró su frecuencia de polígonos (Fig. 17). Comprobamos si la distribución de polígonos de dMWP estaba comportándose como los diagramas de la vía CVT. En este caso, observamos una pequeña desviación de la media de los valores para las imágenes: dMWP presentaba un bajo porcentaje de células con 7 lados en comparación con los diagramas CVT (Fig. 13D y Tabla 5).

En paralelo, analizamos otro tejido donde las propiedades biofísicas de las células eran claramente no fisiológicas (BNA, Materiales y Métodos 1.2). Las atrofas neurógenas son una gran familia de enfermedades neuromusculares que afectan a las neuronas motoras que inervan los músculos. Una característica particular de esta patología es la aparición dispersa de fibras atroficas que son más pequeñas y alargadas (Fig. 16E,F). Esto es consecuencia de la acción de fibras sanas que hipertrofian y aplican presión sobre fibras atroficas que están degenerando (Brazis, 2011; Finsterer y cols., 2011; Mastaglia, 1992). En consecuencia, el histograma de la distribución de área de las células mostró un aumento de células pequeñas en comparación con las imágenes de BCA (Fig. 16G y Tabla 7). Aunque la distribución de polígonos resultante era similar al diagrama 2, presentaba una diferencia significativa debido a un aumento inusual de heptágonos lo cual hizo que también se desviara de la vía CVT (Fig. 13D y Tabla 5).

2.5. Propiedades biofísicas que dirigen la organización de los tejidos naturales empaquetados

Hemos investigado que tipo de alteraciones biofísicas de las células podrían provocar que los tejidos naturales se desvíen de la vía CVT. Siguiendo con la comparación entre los tejidos empaquetados y los diagramas de Voronoi. Los tejidos biológicos no emergen de semillas, pero sí de células que crecen hasta alcanzar un determinado volumen estable ("target area" o "resting volume" en inglés). Esto da a las células una presión interna que provoca que las células se empujen mutuamente, formando un tejido empaquetado de células convexas (Hayashi y Carthew, 2004; Classen y cols., 2005). Teorizamos que en las dos condiciones no fisiológicas que examinamos, el balance original de células con volúmenes estables podría estar comprometido. En el caso de la ausencia de miosina II algunas células podrían continuar creciendo sin desarrollar la citocinesis, intentando alcanzar por tanto un volumen de estabilidad mucho mayor. En el caso de las fibras atroficas, el proceso de degeneración podría reducir su volumen de estabilidad, y de ahí que las células estén deformadas por las fibras contiguas normales.

Medir la presión interna o el volumen de estabilidad de las células en tejidos *in vivo*, no es factible con las técnicas actuales. Por ello, empleamos una aproximación clásica de pérdida de función mediante un modelado *in silico*. Analizamos grupos de células con un volumen de estabilidad no balanceado dentro del tejido. Para realizarlo, desarrollamos diferentes tipos de simulaciones por ordenador usando un modelo de vértices (Farhadifar y cols., 2007; Mao Y. y cols., 2011).

El modelo de vértices es un método de simulación que ha sido utilizado para investigar los procesos físicos que se producen en las células en distintos tipos de tejidos empaquetados (Introducción 4.2). En nuestro estudio utilizamos una ecuación que refleja la energía total del tejido (Farhadifar y cols., 2007) donde las células se someten a deformaciones elásticas, cambios volumétricos y otros movimientos debidos a la adhesión a otras células. Los parámetros que miden la deformación del área son el coeficiente de elasticidad, el área ideal de la célula y el área actual. Para medir el nivel de tensión y adhesión entre las células se define el coeficiente de tensión lineal. Para medir la capacidad de una célula de cambiar de forma se define el coeficiente de contractibilidad (Materiales y métodos 4.2).

Se procedió con la simulación computacional de los eventos biofísicos complejos que ocurren en un tejido proliferativo que presenta mutación en miosina II. Para reproducir un tejido proliferativo se simuló la mitosis. En estas simulaciones, durante el crecimiento cada célula incrementa su área ideal hasta alcanzar el doble de su tamaño original,

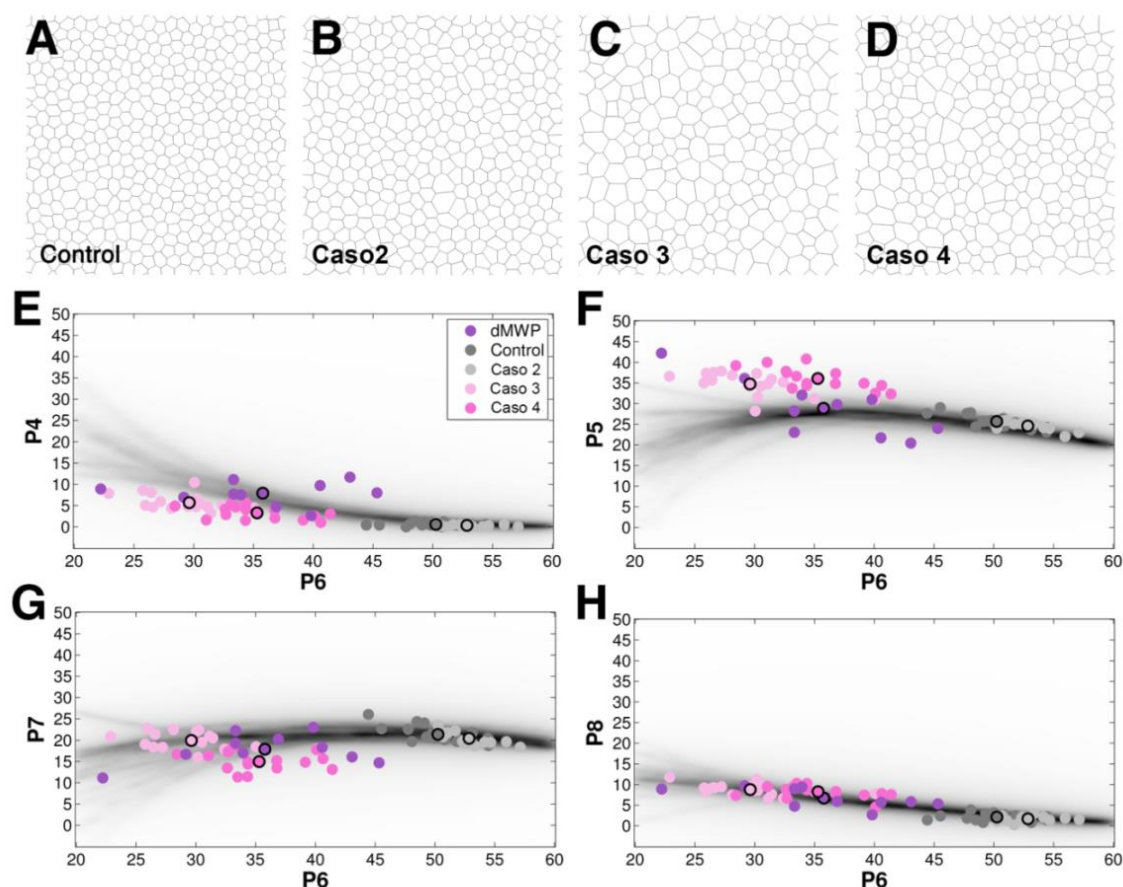


Fig. 18. Alteraciones biofísicas en tejidos proliferativos usando simulación por ordenador. **A)** Diagrama resultante de simular un modelo de vértices que incluye la proliferación celular. **B)** Diagrama resultante de simular un modelo de vértices que incluye la proliferación celular y una reducción heterogénea de la línea de tensión entre las células del tejido. **C)** Diagrama resultante de simular un modelo de vértices con reducción heterogénea de la línea de tensión y una inhibición de la división celular cuando la tensión está por debajo del 30 por ciento del valor inicial. **D)** Simulación similar a (C) con un umbral de 40 por ciento. **E-H)** Gráficos de densidad que muestran la frecuencia de células de cuatro, cinco, siete y ocho lados respectivamente, dependiente del porcentaje de hexágonos (P6) en los diagramas CVT. Los valores de las imágenes individuales de la mutación de miosina (violeta) y los diagramas individuales de las 4 simulaciones (gris claro, gris, rosa claro y rosa) son mostrados en cada gráfica. El valor medio para cada simulación es mostrado con una circunferencia negra.

entonces se divide en 2 células hijas ("Control", Fig. 18A,E-H). Para simular la falta de miosina, en primer lugar, se consideró la pérdida de tensión cortical que ocurre en el tejido mutante, para ello se asignó aleatoriamente una reducción en el parámetro que dictamina la tensión de cada célula. Observamos que esto no cambió sustancialmente la estructura del tejido ("Caso 2", Fig. 18B,E-H). Posteriormente, se incorporó una condición adicional para simular la dificultad para realizar la citocinesis que conlleva la ausencia de miosina II. Las células debían tener un umbral mínimo de tensión para poder dividirse. Si las células no alcanzaban este umbral, éstas continuaban creciendo en tamaño sin ser capaz de dividirse. Una célula sin la capacidad de dividirse se quedaría estancada en la fase mitótica y no comenzaría con un segundo proceso de

división celular. Es por ello, que incrementaba el valor de su área ideal. Se analizaron varios umbrales para intentar simular los datos dMWP. Al igual que ocurrió en los tejidos que presentaban mutación de miosina II donde las imágenes eran muy variables, los resultados obtenidos con cada simulación individual fueron muy diversos. Se analizó con mayor profundidad las simulaciones con un umbral de 30 y 40 por ciento de la tensión control para permitir la división celular debido a que reproducían el amplio rango de nivel medio de hexágonos encontrados en dMWP (“Caso 3” y “Caso 4”, Fig. 18C,D respectivamente). En cada caso hubo una mezcla de imágenes cercanas a la vía CVT con imágenes que se desviaban en diferentes direcciones. En resumen, la alteración del volumen de estabilidad (área ideal en la ecuación) fue crítico para producir la diversidad de resultados y la desviación de la predicción de la vía CVT. Esto rompió el equilibrio original de fuerzas del tejido en diferentes sentidos, permitiendo al tejido adoptar diferentes topologías. Debido al hecho que la alteración única de la tensión lineal era insuficiente para alterar el patrón de empaquetamiento celular de la vía CVT (Fig. 18B), pensamos que el volumen de estabilidad de las células es el principal factor responsable de la ruptura de la vía CVT.

Con el objetivo de esclarecer la importancia del volumen de estabilidad de la célula en el empaquetamiento de tejidos, intentamos simular tejidos de músculos con atrofia. Para ello, comparamos una simulación control donde todas las células presentan la misma área ideal (“Simulación Control”, Fig. 19A) con el efecto de tener una baja área ideal en ciertas células (“Simulación Atrofia”, Fig. 19B) (hay que precisar como detalle técnico que también se incrementó el efecto de adhesión célula a célula para prevenir la desaparición de células “inestables” en la simulación, Materiales y Métodos 4.2). Esto produjo células pequeñas y alargadas (altamente anisotrópicas). Estas células representan a las células afectadas presentes en el tejido BNA. Adicionalmente, comparamos con una simulación donde las células afectadas presentaban las mismas características de contractibilidad y adhesión que la simulación previa pero que presentaban un área equiparable a las células control (“Area Ideal=1”, Fig. 19C). Con el objetivo de entender cómo se organizan estas células afectadas, las comparamos con el 10 por ciento de las células más pequeñas tanto del tejido control (“BCA 10%”) como de las simulaciones control sin proliferación (“Sim Control 10%”). Comprobamos como al tratarse de células pequeñas la mayoría de ellas van a tender a tener un menor número de vecinos en los tejidos control mientras que en la simulación de atrofia las células afectadas presentaron en general un mayor número de vecinos (Fig. 19D y Fig. 20). Nuestros análisis demostraron que esta diferencia en la distribución de vecinos era debido a la alta proporción de células de 7 lados entre las pequeñas fibras atróficas. (Fig. 19D, BNA Afectadas y Fig. 20). Realizamos un análisis más detallado de las

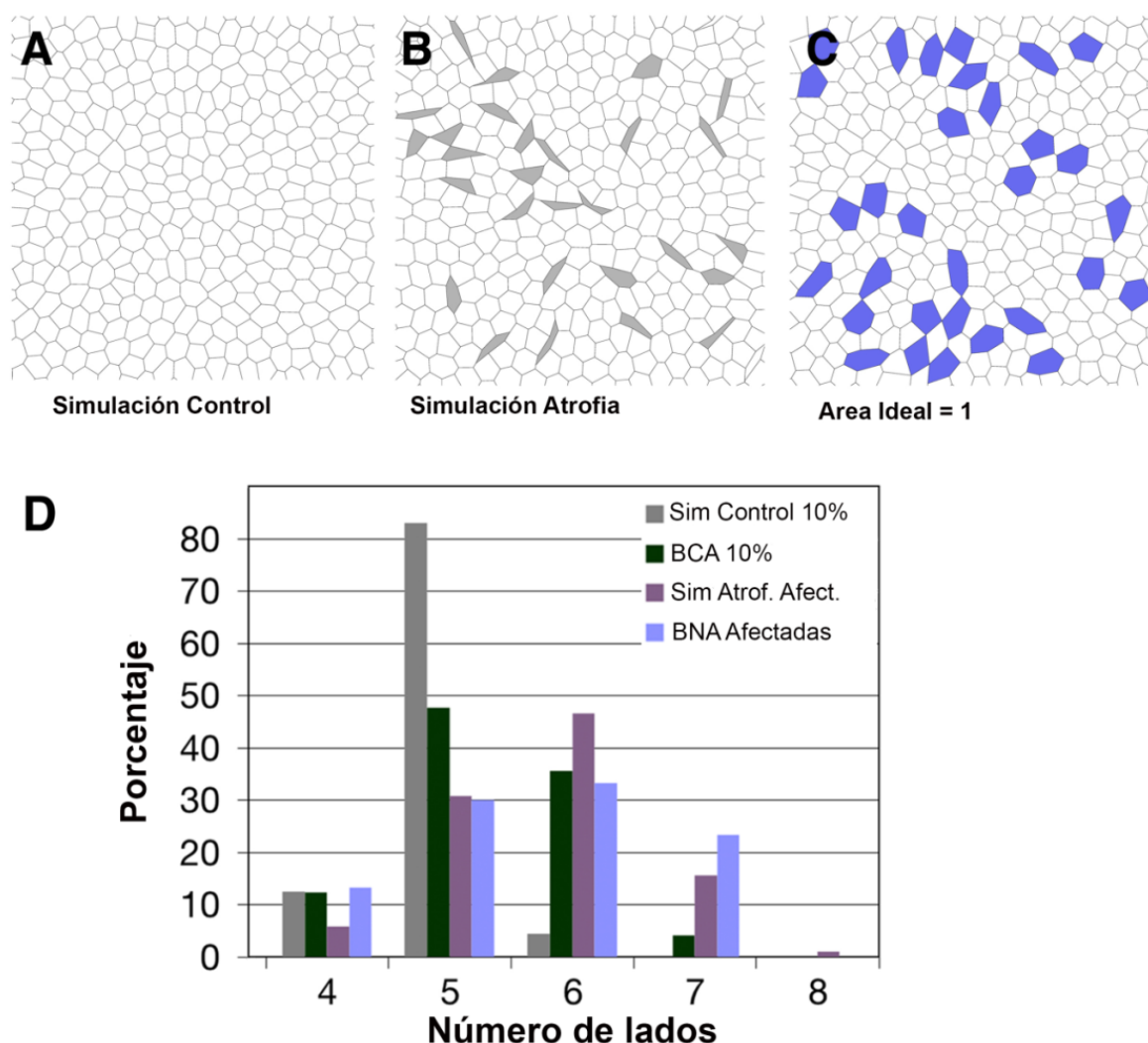


Fig. 19. Simulación por ordenador muestra que homogeneidad en el volumen de estabilidad regula las constricciones del tejido. **A)** Diagrama resultante de simular parámetros de contractibilidad, tensión lineal y área ideal homogéneos usando un modelo de vértices. **B)** Diagrama resultante de simular un modelo de vértices donde se redujo la línea de tensión y el área ideal en el diez por ciento de las células (gris). **C)** Diagrama control donde solamente se redujo la línea de tensión en el 10% de las células. **D)** Comparación de la distribución de polígonos contabilizando el 10% de las células más pequeñas en BCA y en la simulación control (A) y empleando las células enfermas de BNA y de la simulación de atrofia (células grises en B).

células atróficas tanto para cada una de las simulaciones que presentaron células afectadas como para el tejido natural BNA. Comparamos la distribución de vecinos entre las células más pequeñas y las células consideradas como afectadas (Fig. 20). Analizamos en la simulación con área ideal 1 el comportamiento de las células afectadas ("Área Ideal=1 Afectadas") y del 10% de las células más pequeñas ("Área Ideal=1 10%"). Comprobamos que las células más pequeñas tendían a presentar menos números de vecinos que las células afectadas, indicativo que las células afectadas en la simulación área ideal estaban compuestas por células de mayor área. En cambio, en la

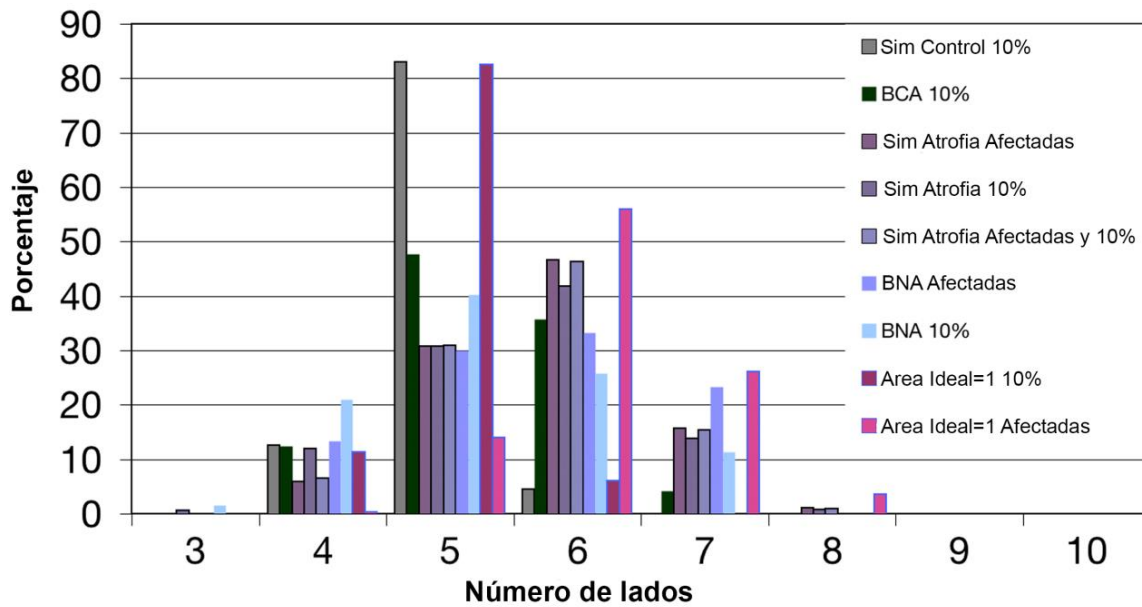


Fig. 20. Distribución de polígonos de diferentes subconjuntos de células provenientes de muestras de músculos y simulaciones. A) Comparación de distribución de polígonos de diferentes subconjuntos de células: Células atroficas de imágenes BNA, 10% de células más pequeñas provenientes de BCA y las 3 simulaciones, y las células “afectadas” seleccionadas con diferentes parámetros para la simulación de atrofia y la simulación con área ideal =1. “Simulación control 10%,” “Simulación Area Ideal=1 10%,” “BCA 10%,” “BNA 10%” y “Simulación de atrofia 10%” presenta un porcentaje más alto de pentágonos que de hexágonos. Por otro lado, cuando se analizaron las “células atroficas” provenientes de “BNA afectadas” y “Simulación atrofia afectadas” se observó un incremento de hexágonos y heptágonos. “Simulación atrofia afectadas,” “Simulación atrofia 10%” y “Simulación atrofia afectadas y 10%” presentan una distribución de polígonos similar debido a que en la simulación muchas de las células afectadas están incluidas dentro del 10% de células más pequeñas.

simulación de atrofia vimos que compartían una distribución similar entre las células afectadas (“Sim Atrofia Afectadas”) y el 10% de células más pequeñas (“Sim Atrofia 10%”). Este hecho es más equiparable a lo que ocurre al comparar las distribuciones de BNA de células afectadas (“BNA Afectadas”) con el 10% de células más pequeñas en BNA (“BNA 10%”). Mediante este análisis de células atroficas, observamos que estaban rompiendo una restricción fundamental: La relación directa (lineal o no) de que células grandes presentan mayor número de lados, y viceversa.

Discusión

Esta tesis versa sobre la caracterización y comprensión del establecimiento y mantenimiento de la organización celular mediante el análisis de imágenes biológicas y biomédicas. Para ello, se recurre al tratamiento avanzado de imágenes para extraer información morfológica de cada célula y al empleo de grafos para capturar información relevante sobre cómo se organizan dichas células en el tejido.

En la primera parte de la tesis hemos estudiado de forma cuantitativa la morfogénesis de epitelios proliferativos, empleando imágenes del disco del ala de *Drosophila* como modelo. Para investigar el efecto que tiene en la organización los mecanismos de proliferación de los tejidos se han aplicado diversos métodos para comparar de forma exhaustiva dos estadios de desarrollo muy cercanos.

A partir de ese punto hemos comenzado un nuevo abordaje en el que pretendemos encontrar las restricciones que nos permitan definir cómo las células se organizan para formar el tejido. Para ello hemos comparado diversos tejidos con distinto origen y provenientes de diferentes organismos con un nuevo concepto geométrico desarrollado en este trabajo, la vía CVT.

1. Influencia de la proliferación en la organización de tejidos epiteliales

El análisis de pequeños cambios globales durante el desarrollo del epitelio es clave para comprender los mecanismos que coordinan crecimiento y morfogénesis. Hemos usado la ampliamente estudiada estructura epitelial del disco de ala de *Drosophila* como modelo para realizar nuestro estudio. Comparamos propiedades tanto geométricas como topológicas de las fases del tercer estadio de la larva y el inicio de la prepupa, dos puntos en el tiempo del desarrollo epitelial que no presentan una organización reconocible en el tejido (Fig. 6A). Al final del desarrollo, el ala se organiza en forma de un empaquetamiento de células hexagonales. Esto ha sido considerado como dependiente de los genes de polaridad celular planar que actúan durante las etapas de pupa (Gho y cols., 1999). De ahí que, la fase de proliferación en el desarrollo de la larva haya sido clásicamente vista como un estadio activo de crecimiento con escasa coordinación en la reorganización, más allá del mantenimiento de una distribución poligonal estereotipada (Gibson y cols., 2006). Nuestros resultados sugieren la existencia de pequeñas pero significativas diferencias topológicas entre el tercer estadio larvario y el estadio de prepupa, marcando el comienzo del reordenamiento del tejido en la fase proliferativa.

Comenzamos buscando diferencias a nivel de células entre distintos tejidos. Usamos un estudio de correlación para comparar la similitud entre todas las células que integran nuestros 2 grupos de imágenes. Este análisis produce un claro resultado:

los nodos (células) no están agrupados dependiendo de la fase de desarrollo (tercer estadio o prepupa) (Fig. 6B). Por el contrario, la distribución en el grafo está completamente relacionada con el número de lados de las células: células de 6 lados se enlazan preferentemente a otros hexágonos, los pentágonos forman agrupaciones y también se propician grupos de heptágonos (Fig. 6C). Este experimento presenta una clara conclusión: los valores de las 14 características usadas en la correlación (Tabla 1) presentan una dependencia con el número de lados de las células (aunque el número de lados no sea una de las 14 características utilizadas). Esta dependencia no afecta solo a las características geométricas sino también a las características topológicas de células individuales.

Por otro lado, encontramos que las distribuciones del porcentaje de polígonos de dWL y dWP son significativamente diferentes. Combinando las consecuencias de estos dos resultados, se puede deducir que si el número de lados de una célula influencia sus características topológicas locales, una alteración significativa en la distribución de polígonos puede ser convertida a una transformación de la organización epitelial. En otras palabras, cambios en la prevalencia de un polígono sobre otro puede llevar a la modificación de toda la organización en el plano del epitelio (Fig. 21).

La pregunta subyacente es si solo una diferencia en la distribución de polígonos puede explicar cambios en la organización epitelial. Se comprobaron 4000 combinaciones de grupos arbitrarios de imágenes para concluir que este no es el caso. Además, no existe correlación entre las distribuciones de polígonos de ambos grupos y la diferenciación en organización capturada por el PCA. Consecuentemente, solo la combinación que tiene sentido biológico es significativa en ambos tests y mejor que las otras 4000 combinaciones (Fig. 11). Por esta razón, se infiere que durante el desarrollo, los cambios en la distribución de polígonos que ocurren entre la tercera fase de la larva y la fase temprana de prepupa están altamente coordinados. Esto posibilita la aparición de cambios importantes en la organización de las células que ahora pueden ser capturados con nuestro nuevo método basado en grafos.

Estos cambios en la organización se pueden describir mediante las características relevantes usadas para diferenciar dWL de muestras de dWP. Aquellas más relevantes son: Media del eje mayor, Media de relación de envolvente convexa de vecinos y Media de intermediación.

Estas 3 características son elegidas según el orden en las que aparecen en el paso de selección de características y su peso en el PCA. Es interesante mencionar que estas 3 características también fueron utilizadas en el experimento de correlación siendo incapaces de distinguir entre tejidos a nivel celular.

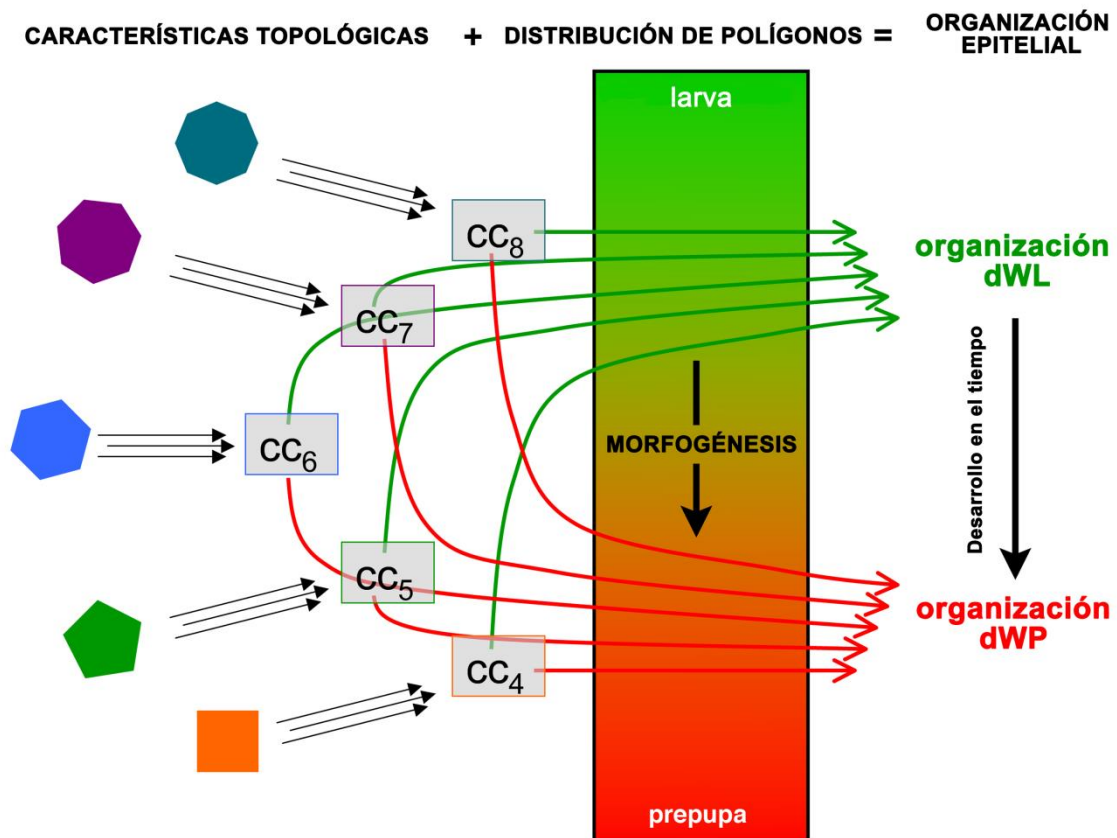


Fig. 21. Modelo regulatorio de la organización del tejido epitelial durante el final de la fase larvaria proliferativa. Las señales morfogenéticas en el disco de ala provocan un cambio en la distribución de polígonos del tejido epitelial del disco imaginal entre los estadios de larva (verde) y prepupa (rojo). El número de lados de las células epiteliales impone restricciones en la geometría local que se transmiten a la organización de todo el tejido.

¿Cuál es el sentido biológico de estas características? Su definición puede ayudar a explicar los cambios durante el intervalo de 24 horas entre los dos estadios de desarrollo.

En el caso de las características geométricas tales como la media del eje mayor, es fácil de interpretar que células de dWP crecen incrementando el valor de su eje mayor (lo mismo ocurre con los valores del eje menor y del área). El caso de la media de relación de envolvente convexa de vecinos es más complicado. La envolvente convexa se define como el objeto convexo más pequeño que contiene la forma de la célula. Sin embargo, para tener una referencia con respecto a la célula, se computa el valor como el cociente entre el área y la envolvente convexa del área celular. De ahí que un valor cercano a 1 indica que la célula presenta una forma convexa con lados suaves y rectos. Un incremento en las irregularidades del perímetro de las células decrementa este valor. Nuestros datos indican que el contorno de las células dWP es más regular puesto que en este caso los valores de la media de la envolvente convexa y la media de relación de envolvente convexa de vecinos son cercanos a 1. Esto se

puede interpretar como que en las imágenes dWP, las células son más similares a sus vecinos en términos de regularidad de contorno, siendo una de las características más importantes para discriminar entre dWL y dWP. En términos biológicos, esta característica puede expresar los efectos del estrés mecánico que sufren el contorno de las células. Las concentraciones de células de prepupa presentarán, tensiones más altas inducidas por los movimientos morfogénicos que ocurren en este estadio (Aldaz y cols., 2010). La tercera característica más discriminante es la media de la intermediación definida como la fracción de todos los caminos más cortos en la network que contiene un nodo dado. Nodos con alto valor de intermediación participan en un número mayor de caminos cortos, ellos son usualmente llamados “hubs” (nodo con múltiples conexiones) (Freeman, 1979; Newman, 2005). dWL presenta un mayor número de estos “hubs” combinado con otras células con un valor más bajo. Esto nos sugiere que el tejido es más heterogéneo en términos de conectividad de sus nodos. Por otro lado, las muestras de prepupa indican un valor pequeño en la media de la intermediación y en su desviación estándar. Estos resultados, trasladados a nuestras imágenes, indican un incremento de la homogeneidad del tejido con respecto a dWL. Las 2 características restantes presentan un peso bajo en el PCA y sirven para complementar las 3 características previas que poseen un mayor peso a la hora de identificar diferencias entre ambos grupos. La eficiencia y la desviación de la fuerza estarían también relacionadas con el incremento de la homogeneidad y del tamaño respectivamente.

Para resumir, las características relevantes capaces de separar imágenes de dWL y dWP indican que el enriquecimiento de células de 6 lados comienza cuando el disco está todavía creciendo. Un aumento del tamaño de las células, un suavizado de su contorno y un incremento de la homogeneidad de todo el epitelio fomenta la organización de células hexagonales que aparecerá claramente en el estadio de pupa tardío (Classen y cols., 2005). Nuestro nuevo método de análisis marca ahora el comienzo del empaquetamiento hexagonal, durante la fase proliferativa del desarrollo de la larva.

2. Restricciones físicas de las células como mecanismo regulatorio de la organización

En esta tesis, hemos profundizado también en como se organizan los tejidos empaquetados presentes en la naturaleza más allá de si proliferan o no. Para ello, hemos realizado un estudio amplio de diversos tipos de tejidos empaquetados tanto proliferativos como no proliferativos. La evaluación de las muestras de tejido ha sido

realizada utilizando la distribución de polígonos como una primera medida de organización (Tabla 4). Antes de nuestro estudio, estaba establecida la existencia de una distribución de polígonos que se conserva en todos los epitelios proliferativos de metazoos (Gibson y cols., 2006; Aegerter-Wilmsen y cols., 2010; Nagpal y cols., 2008; Patel y cols., 2009). Primero, nuestro estudio revela la existencia de tejidos proliferativos que presentan dicha distribución estereotipada (dWL, dWP), tejidos proliferativos que no presentan la distribución estereotipada (cNT) y tejidos no proliferativos que presentan la distribución estereotipada (BCA). A la vez realizamos un abordaje geométrico usando los diagramas de Voronoi y encontramos que un incremento de la homogeneidad del tamaño de las células permitirá alcanzar un alto porcentaje de hexágonos, pentágonos y heptágonos. Esto explica la distribución de polígonos estereotipada publicada previamente, donde los tejidos analizados eran muy homogéneos en términos del área de sus células (Gibson y cols., 2006). Sin embargo, estos autores no contemplaron otros tejidos naturales que presentan diversidad de áreas en su composición como ocurre en el caso de cNT. La evaluación conjunta de todos estos resultados nos lleva a considerar que a diferencia de lo que estaba establecido en el campo, la distribución de polígonos estereotipada no es dependiente de la proliferación celular.

La simulación geométrica del tejido mediante los diagramas de Voronoi nos permite tener un modelo genérico que nos ayuda a comprender la emergencia de organización de tejidos empaquetados. Encontramos que los tejidos empaquetados en condiciones normales fisiológicas presentan distintas distribuciones de polígonos y que todas se corresponden con alguno de los tejidos geométricos de la vía CVT (Fig. 22) debido principalmente a que Voronoi refleja alguna de las propiedades físicas de las células. Adicionalmente tenemos que los cambios fisiológicos que se producen de forma balanceada en la contractilidad y/o adhesión están incrementando los niveles de heterogeneidad en el área de la células, modificando pero manteniendo, el tejido a lo largo de la vía CVT (Fig. 22). Por tanto, nuestros resultados apoyan la existencia de unas restricciones físicas que afectan a cualquier tejido natural o teselación matemática compuesta por polígonos convexos. Esto hace que no puedan presentar distribuciones de polígonos infinitas, sino limitadas a las descritas por la vía CVT.

Hemos buscado las bases biofísicas que hacen que los tejidos tengan restringida la forma de organizarse. Para ello analizamos las propiedades de las muestras de tejidos que presentan alteraciones no fisiológicas de las propiedades biofísicas de las células. Estudiamos condiciones, tales como la disminución de miosina II en el tejido dMWP o la atrofia de las fibras musculares en el tejido BNA. Ambos tejidos no son capaces de reproducir completamente la organización tisular que dicta la vía CVT (Fig. 22). Estos

casos tienen en común que el balance de fuerzas entre sus células está comprometido. En el caso de la ausencia de miosina II algunas células podrían continuar creciendo y entrar en división sin realizar la citocinesis, teniendo por tanto un volumen de estabilidad mucho mayor que otras células de alrededor. En el caso de las fibras atroficas, el proceso de degeneración podría reducir el volumen con el cual la célula alcanza la estabilidad, y de ahí que las células estén deformadas por la presión de las fibras adyacentes normales.

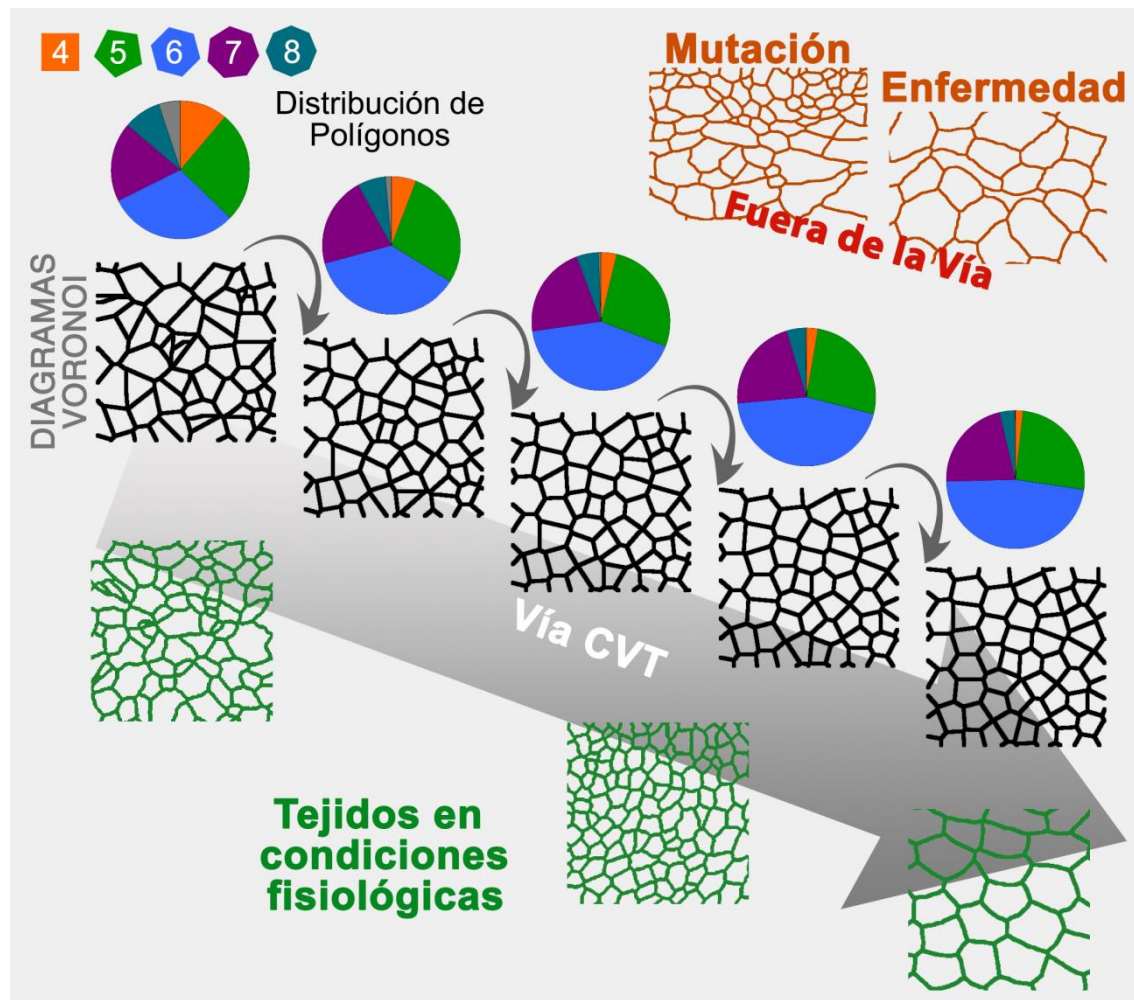


Fig. 22. Esquema conceptual de la organización de tejidos empaquetados. Los tejidos en condiciones fisiológicas (verde) presentan unas limitaciones físicas que hacen que sus distribuciones de polígonos estén limitadas a la vía CVT (establecida por diagramas de Voronoi cada vez mas homogéneos). Tejidos bajo condiciones no fisiológicas (mutación o enfermedad, naranja) salen fuera de la vía CVT.

Ante la imposibilidad de cuantificar los mecanismos biofísicos que regulan el comportamiento de estas células alteradas, intentamos modelar sus características mediante el sistema de modelado de vértices en los tejidos dMWP y BNA.

Para modelar dMWP ejecutamos tres simulaciones ("Caso II", "Caso III" y "Caso IV") considerando el efecto de la tensión celular en el comportamiento celular de un

tejido proliferativo. Analizando la organización celular del “Caso II” comprobamos como a pesar de una reducción de la tensión celular la organización no se ve sustancialmente afectada. En las simulaciones “Caso III” y “Caso IV” conseguimos reproducir la amplia variedad de distribuciones de polígonos que presentan las imágenes de la categoría dMWP. En estas dos últimas simulaciones interferíamos con la división celular y hacíamos crecer el área ideal de algunas células al azar (debido a la aleatorización de los valores de la tensión). Deducimos que la alteración del volumen de estabilidad, representado en la simulación por el área ideal, es la principal responsable de la ruptura de fuerza que regula la topología del tejido provocando que se desvíe de la vía CVT.

Para modelar BNA realizamos dos simulaciones de células atróficas en tejidos no proliferativos (“Simulación Atrofia” y “Area Ideal = 1”). Estudiando la organización de las células en BNA hemos comprobado que muchas de las células pequeñas tienen un número anormalmente alto de vecinos. Este hecho rompe con las leyes geométricas de tejidos empaquetados que enuncia que células pequeñas tienden a tener menor número de vecinos. Analizamos la organización tanto de células afectadas como las células con un tamaño más pequeño en las distintas simulaciones. Encontramos que las células afectadas en “Simulación Atrofia” se comportan como las de BNA al presentar también células pequeñas con un alto número de vecinos. Estas células simuladas fueron obtenidas alterando principalmente el valor del área ideal (y variando ligeramente la adhesión celular). Sin embargo, la simulación “Area Ideal = 1” (obtenida alterando únicamente la adhesión celular) no origina células como las de BNA. Estos experimentos indican que el volumen de estabilidad (definido a través del área ideal) es el principal parámetro responsable de la alteración de la distribución de polígonos en todo el tejido y causante de la ruptura de BNA con la vía CVT.

En resumen, hemos encontrado que los tejidos naturales empaquetados no se pueden ordenar libremente en infinitas organizaciones. En su lugar, éstas están limitadas a unas ciertas topologías las cuales hemos sido capaces de definir mediante su comparación con los diagramas de Voronoi. Nuestros resultados muestran que esto es debido a restricciones físicas inducidas por el balance normal de las fuerzas fisiológicas presentes entre las células o las fibras. Los diagramas de Voronoi y los tejidos naturales comparten esta restricción, es por ello que todos los tejidos bajo estas condiciones fisiológicas converjan con la vía CVT. En condiciones no fisiológicas, tales como en una enfermedad o una mutación que afecte al citoesqueleto, este balance original de fuerzas en el volumen de estabilidad de las células puede romperse, lo cual lleva a los tejidos a desviarse de la escala marcada por la vía CVT.

Las células con formas anisotrópicas, las cuales pueden afectar a la correlación entre el área y el número de lados de una célula (Kim y cols., 2014), pueden ser un factor que rompa las restricciones físicas de estas teselaciones. Los datos obtenidos a partir del tejido de músculo atrófico junto con nuestros experimentos *in silico* sugieren que el volumen de estabilidad de la célula, el cual crea una presión celular interna, es el principal componente biofísico que establece la restricción física para los tejidos empaquetados. Cualquier desviación patológica del volumen de estabilidad fisiológico llevará a la ruptura de esta restricción y originará un nuevo tejido empaquetado cuya distribución de polígonos podrá ser diferente del camino CVT.

Una importante vía para futuras investigaciones sería probar si esto es cierto para los tejidos en diferenciación y si las desviaciones de los diagramas de CVT pueden detectar el cambio a un nuevo tipo celular. Esto también podría representar un nuevo método de análisis de imagen para el diagnóstico precoz del inicio de enfermedad.

Conclusiones

Mediante la realización de esta tesis se ha perseguido profundizar tanto en la morfogénesis como el mantenimiento de la organización celular en tejidos empaquetados. Los resultados obtenidos permiten concluir que:

- 1) Hemos desarrollado un método basado en la Teoría de grafos que permite distinguir imágenes epiteliales muy similares de un mismo organismo, identificando características relevantes que permiten encontrar diferencias en su organización epitelial durante el desarrollo.
- 2) En epitelios proliferativos el número de lados de las células impone restricciones en la geometría local que afectan a la organización general del tejido.
- 3) La distribución estereotipada de polígonos en epitelios proliferativos de metazoos no es dependiente de la proliferación celular. Es debida a que las células que los componen presentan una alta homogeneidad de tamaños.
- 4) La vía CVT establece una escala que puede ser utilizada para cuantificar la organización que adquiere un tejido natural empaquetado.
- 5) Los diagramas geométricos de Voronoi pueden reflejar propiedades físicas de las células y los cambios fisiológicos que se producen en los tejidos.
- 6) Las distribuciones de polígonos de los tejidos en condiciones fisiológicas normales están limitadas a ciertas topologías que coinciden con la escala marcada por la vía CVT. En condiciones no fisiológicas que rompen el balance de fuerzas entre las células de los tejidos se produce una desviación del comportamiento descrito por la vía CVT.
- 7) El volumen de estabilidad de la célula es el principal componente biofísico que establece la restricción física que rige la organización celular en los tejidos empaquetados.
- 8) La vía CVT se postula como una escala que puede ser adaptada para diferenciar distintos estadios en un tejido empaquetado. Las desviaciones en la organización respecto a la vía CVT pueden tener relevancia diagnóstica para detectar condiciones no fisiológicas de los tejidos.

Bibliografía

- Aboav, D. A. (1970). The arrangement of Grains in a Polycrystal. *Metallography*(3), 383-390.
- Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A. y Gocayne, J. D. (2000). The genome Sequence of *Drosophila melanogaster*. *Science*, 287, 2185-2195.
- Aegerter-Wilmsen, T., Smith, A. C., Christen, A. J., Aegerter, C. M., Hafen, E. y Basler, K. (2010). Exploring the effects of mechanical feedback on epithelial topology. *Development*(137), 499-506.
- Aldaz, S., Escudero, L. M. y Freeman, M. (2010). Live imaging of *Drosophila* imaginal disc development. *PNAS*, 107, 14217-14222.
- Aldaz, S., Escudero, L. M. y Freeman, M. (2013). Dual role of myosin II during *Drosophila* imaginal disc metamorphosis. *Nature Communications*, 1761.
- Axelrod, J. D. (2006). Cell shape in proliferating epithelia: a multifaceted problem. *Cell*, 126, 643-645.
- Barret, K., Leptin, M., y Settleman, J. (1997). The rho gtpase and a putative rhogef mediate a signaling pathway for the cell shape changes in *Drosophila* gastrulation. *Cell*, 91, 905-915.
- Bertet, C., Sulak, L. y Lecuit, T. (2004). Myosin-dependent junction remodelling controls planar. *Nature*, 429, 667-671.
- Boots, B. N. (1980). Packing polygons: some empirical evidence. *Canadian Geographer*, 406-411.
- Brazis, P. W. (2011). General Principles of Neurological Localization. En *Localization in clinical neurology*. Philadelphia: Lippincott Williams & Wilkins.
- Brodlan, G. W., Viens, D. y Veldhuis, J. H. (2007). A new cell-based FE model for the mechanics of embryonic epithelia. *Comput. Methods Biomech. Biomed. Engin.*, 10, 121-128.
- Brodland, G. W. (2004). Computational modeling of cell sorting, tissue engulfment, and related phenomena: a review. *Appl. Mech. Rev.*, 57, 47-76.
- Brodland, G. W., Chen, D. I. y Veldhuis, J. H. (2006). A cell-based constitutive model for embryonic epithelia and other planar aggregates of biological cells. *Int. J. Plast.*, 22, 965-995.
- Chichilnisky, E. J. (1986). A mathematical model of pattern formation. *J Theor Biol*, 123, 81-101.
- Chiu, S. (1995). Aboav-Weaire's and Lewis' laws—A review. *Materials Characterization*, 34, 149-165.

- Chiu, S. N. (1995). Aboav-Weak's and Lewis' Laws-A Review. *Materials Characterization*(34), 149-165.
- Classen, A. K., Anderson, K. I., Marois, E. y Eaton, S. (2005). Hexagonal Packing of Drosophila Wing Epithelial Cells by the Planar Cell Polarity Pathway. *Developmental Cell*, 9, 805-817.
- Conti, M. A., Kawamoto, S. y Adelstein, R. S. (2008). Non-Muscle Myosin II. En L. M. Coluccio, *Myosins: A Superfamily of Molecular Motors* (págs. 223-264). Dordrecht, The Netherlands.: Springer.
- Costa, L. D., Rodriguez, F. A., Travieso, G. y Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. *Advances in physics*(56), 167-242.
- Costa, L. y Cesar, R. M. (2009). *Shape Classification and Analysis: Theory and Practise* (pág. 680). CRC Press.
- Costa, M., Wilson, E. T. y Wieschaus, E. (1994). A putative cell signal encoded by the folded gastrulation. *Cell*, 76, 1075-1089.
- Cui, D. (2010). *Atlas of histology with functional & clinical correlation*. Wolters Kluwer Health.
- Du, Q., Faber, V. y Gunzburger, M. (1999). Centroidal Voronoi Tessellations: Applications and Algorithms. *SIAM review*(41), 637-676.
- Dubowitz, V. y Sewry, C. A. (2007). *Muscle Biopsy: A practical approach*. Saunders Elsevier, ed. Third edition.
- Eaton, S. y Jülicher, F. (2011). Cell flow and tissue polarity patterns. *Curr Opin Genet Dev*, 747-752.
- Escudero, L. M., Bischoff, M. y Freeman, M. (2007). Myosin II regulates complex cellular arrangement and epithelial architecture in Drosophila. *Dev Cell*(13), 717-729.
- Escudero, L. M., Costa, L. D., Kicheva, A., Briscoe, J., Freeman, M. y Babu, M. M. (2011). Epithelial organisation revealed by a network of cellular contacts. *Nat Commun*(2), 526.
- Fagiolo, G. (2007). Clustering in Complex Directed Networks. *Physical Review E*, 76, 026107.
- Farhadifar, R., Roper, J. C., Aigouy, B., Eaton, S. y Jülicher, F. (2007). The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Curr Biol*(17), 2095-2104.
- Finsterer, J., Papic, L. y Auer-Grumbach, M. (2011). Motor neuron, nerve, and neuromuscular junction disease. *Curr Opin Neurol*(24), 469-474.

- Fletcher, A. G., Osterfield, M., Baker, R. E y Shvartsman, S. Y. (2014). Vertex Models of Epithelial Morphogenesis. *Biophys J.*, 106, 2291-2304.
- Frade, J. M. (2002). Interkinetic nuclear movement in the vertebrate neuroepithelium: encounters with an old acquaintance. *Prog Brain Res*(136), 67-71.
- Franke, J. D., Montague, R. A. y Kiehart, D. P. (2010). Nonmuscle myosin II is required for cell proliferation, cell sheet adhesion and wing hair morphology during wing morphogenesis. *Developmental Biology*, 345, 117-132.
- Fukunaga, K. (1990). Introduction to statistical pattern recognition. *Academic Press*.
- Gervois, A., Troadec, J. P. y Lemaitre, J. (1992). Universal properties of Voronoi tessellations of hard discs. *J Phys A: Math Gen*(25), 6169-6177.
- Gho, M., Bellaïche, Y. y Schweisguth, F. (1999). Revisiting the *Drosophila* microchaete lineage: a novel intrinsically asymmetric cell division generates a glial cell. *Development*, 126, 3573-3584.
- Gibson, M. C. y Schubiger, G. (2001). *Drosophila* peripodial cells, more than meets the eye? *Bioessays*, 23, 691.
- Gibson, M. C., Patel, A. B., Nagpal, R. y Perrimon, N. (2006). The emergence of geometric order in proliferating metazoan epithelia. *Nature*(442), 1038-1041.
- Gibson, W. T. y Gibson, M. C. (2009). Current Topics in Developmental Biology. *Curr Top Dev Biol.*, 89, 87-114.
- Gibson, W. T., Veldhuis, J. H., Rubinstein, B., Cartwright, H. N., Perrimon, N., Brodland, G. W., Radhika, N. y Gibson, M. C. (2011). Control of the mitotic cleavage plane by local epithelial topology. *Cell*, 144, 427-438.
- Gomez-Alvarez, R. P. (2007). *Citología e histología vegetal y animal (4ª edición)*. MCGRAW-HILL.
- Good, B. H., de Montjoye, Y.-A. y Clauset, A. (2010). Performance of modulatory maximization in practical contexts. *Physical Review E*, 046106.
- Haigo, S. L., Hidebrand, J. D. y Harland, R. M. (2003). Shroom Induces Apical Constriction and Is Required for Hinge-point Formation during Neural Tube Closure. *Current Biology*, 13, 2125-2137.
- Hardin, J., y Walston, T. (2004). Models of morphogenesis: the mechanisms and mechanics of cell rearrangement. *Curr. Opin. Genet. Dev.*, 14, 399-406.

- Hayashi, T. y Carthew, R. W. (2004). Surface mechanics mediate pattern formation in the developing retina. *Nature*, 431, 647-652.
- Helliwell, T. R. (1999). Muscle: Part 1 - Normal structure and function. *Curr Orthop*, 33-41.
- Hildebrand, J. D. y Soriano, P. (1999). Shroom, a PDZ Domain-Containing Actin-Binding Protein, Is Required for Neural Tube Morphogenesis in Mice. *Cell*, 99, 485-497.
- Hilhorst, H. J. (2006). Planar Voronoi cells: the violation of Aboav's law explained. *J Phys A: Math Gen*(39), 7227-7243.
- Hocevar, A., El Shawish, S. y Ziherl, P. (2010). Morphometry and structure of natural random tilings. *Eur Phys J E Soft Matter*, 33, 369-375.
- Honda, H. (1978). Description of cellular patterns by Dirichlet domains: The two-dimensional case. *J Theor Biol*, 72, 523-543.
- Honda, H. (1983). Geometrical models for cells in tissues. *Int. Rev. Cytol*, 81, 191-248.
- Honda, H. y Eguchi, G. (1980). How much does the cell boundary contract in a monolayered cell sheet? *J. Theor. Biol.*, 84, 575-588.
- Honda, H., Ogita, Y., Higuchi, S. y Kani, K. (1982). Cell movements in a living mammalian tissue: long-term observation of individual cells in wounded corneal endothelia of cats. *J. Morphol.*, 174, 25-39.
- Junqueira, L. C. y Carneiro, J. (2015). *Histología Básica. Texto y Atlas*. Mexico: Médica Panamericana.
- Kim, S., Cai, M. y Hilgenfeldt, S. (2014). Lewis' law revisited: the role of anisotropy in size-topology correlations. *New Journal of Physics*(16).
- Kölsch, V., Seher, T., Fernandez-Ballester, G. J., Serrano, L. y Leptin, M. (2007). Control of *Drosophila* gastrulation by apical localization of adherens junctions and rhogef2. *Science*, 315, 384-386.
- Korn, R. W. y Spalding, R. M. (1973). The geometry of plants epidermal cells. *New Phytol*(72), 1357-1365.
- Latora, V. y Marchiori, M. (2001). Efficient Behaviour of Small-World Networks. *Physical Review letters*, 87, 198701.
- Le Caër, G. y Delannay, R. (1993). The administrative divisions of mainland France as 2D random cellular structures. *J. Physique 1 France*, 1777-1800.
- Lecuit, T. y Le Goff, L. (2007). Orchestrating size and shape during morphogenesis. *Nature*, 450, 189-192.

- Leptin, M. y Grunewald, B. (1990). Cell shape changes during gastrulation in *Drosophila*. *Development*, 110, 73-84.
- Lewis, F. T. (1928). The correlation between cell division and the shapes and sizes of prismatic cells in the epidermis of cucumis. *Anatomic Rec*(38), 341-376.
- Li, L., Alderson, D., Doyle, J. C. y Willinger, W. (2005). Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications. *Internet Mathematics*, 2(4), 431-523.
- Lloyd, S. P. (1982). Least Squares Quantization in PCM. *IEEE Transactions on Information Theory*(28), 129-137.
- Loeb, A. L. (1976). *Space Structures*. London: Addison-Wesley Pub. Co.
- Mao, Y., Tournier, A. L., Bates, P. A., Gale, J. E., Tapon, N. y Thompson, B. J. (2011). Planar polarization of the atypical myosin Dachs orients cell divisions in *Drosophila*. *Genes Dev*, 131-136.
- Mao, Y., Tournier, A. L., Bates, P. A., Gale, J. E., Tapon, N. y Thompson, B. J. (2011). Planar polarization of the atypical myosin Dachs orients cell divisions in *Drosophila*. *Genes Dev*(25), 131-136.
- Mao, Y., Tournier, A. L., Hoppe, A., Kester, L., Thompson, B. J. y Tapon, N. (2013). Differential proliferation rates generate patterns of mechanical tension that orient tissue growth. *EMBO J*(32), 2790-2803.
- Martín, F. A., Herrera, S. C. y Morata, G. (2009). Cell competition, growth and size control in the *Drosophila* wing imaginal disc. *Development*, 136, 3747-3756.
- Martinez-Arias, A. y Stewart, A. (2002). *Molecular Principles of Animal Development*. Oxford University Press.
- Mastaglia, F. L. (1992). En *Skeletal muscle pathology*. London, Churchill Livingstone.
- McClure, K. D. y Schubiger, G. (2005). Developmental analysis and squamous morphogenesis of the peripodial epithelium in *Drosophila* imaginal discs. *Development*, 5033-5042.
- Milan, M., Campuzano, S. y Garcia-Bellido, A. (1996). Cell cycling and patterned cell proliferation in the *Drosophila* wing during metamorphosis. *PNAS*, 93, 11687-11692.
- Miles, R. E. y Maillardet, R. J. (1982). The basic structure of Voronoi and generalized Voronoi polygons. *J Appl*(19), 97-111.
- Nagpal, R., Patel, A. y Gibson, M. C. (2008). Epithelial topology. *BioEssays*(30), 260-266.
- Newman, M. (2002). Assortative Mixing in Networks. *Physical Review Letters*, 208701.

- Parks, S. y Wieschaus, E. (1991). The *Drosophila* gastrulation gene *concertina* encodes a g alphaslike. *Cell*, 64, 447-458.
- Patel, A. B., Gibson, W. T., Gibson, M. C. y Nagpal, R. (2009). Modeling and Inferring Cleavage Patterns in Proliferating Epithelia. *PloS Comput Biol*(5(6)).
- Pette, D. y Staron, R. S. (2000). Myosin isoforms, muscle fiber types, and transitions. *Microscopy Research and Techique*, 50, 500-509.
- Pignol, V., Delannay, R. y Le Caër, G. (1993). Characterization of topological properties of 2D cellular structures by image analysis. *preprint*.
- Pilot, F. y Lecuit, T. (2005). Compartmentalized Morphogenesis in epithelia: From cell to tissue shape. *Dev Dyn*(232), 685-694.
- Rivier, N. y Schliecker, G. D. (1995). The stationary state of epithelia. *Acta Biotheorica*(43), 403-423.
- Rogers, C. A. (1964). *Packing and covering*. London: Cambridge University Press.
- Ross, M. H. y Wojciech, P. (2007). *Histología. Texto y Atlas color con Bioogía Celular y Molecular*. Argentina: Medica Panamericana.
- Sáez, A., Rivas, E., Montero-Sánchez, A., Paradas, C., Acha, B., Pascual, A., Serrano, C. y Escudero, L. M. (2013). Quantifiable diagnosis of muscular dystrophies and neurogenic atrophies through network analysis. *BMC Medicine*(11), 77.
- Sanchez-Gutierrez, D., Sáez, A., Rivas, E., Pascual, A. y Escudero, L. M. (2013). Topological progression in proliferating epithelia is driven by a unique variation in polygon distribution. *PLoS One*(8).
- Sangwoo, K., Muyun, C. y Sascha, H. (2014). Lewis' law revisited: the role of anisotropy in size-topology correlations. *New Journal of Physics*, 16, 1367-2630.
- Shier, D., Butler, J. L. y Lewis, R. (1999). *Hole's Human Anatomy and Physiology* (Octava ed.). McGraw-Hill Companies.
- Simoës, S., Blankenship, J. T., Weitz, O., Farrel, D. L., Tamada, M., Fernandez-Gonzalez, R. y Zallen, J. A. (2010). Rhokinase. *Dev Cell*, 19, 377-388.
- Voronoi, G. F. (1908). Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les paralléloèdres primitifs. *Journal für die reine und angewandte Mathematik*, 134, 198-287.
- Weaire, D. (1974). Some remarks on the arrangement of grains in a polycrystal. *Metallography*(7), 157-160.

- Weire, D. y Rivier, N. (1984). Soap, cells and statistics—random patterns in two dimensions. *Contemporary Physics*, 25, 59-99.
- Weliky, M. y Oster, G. (1990). The mechanical basis of cell rearrangement. I. Epithelial morphogenesis during *Fundulus* epiboly. *Development*, 109, 373-386.
- Weliky, M., Minsuk, S. y Oster, G. (1991). Notochord morphogenesis in *Xenopus laevis*: simulation of cell behavior underlying tissue convergence and extension. *Development*, 1231-1244.
- Wigmore, P. M. y Dunglison, G. F. (1998). The generation of fiber diversity during myogenesis. *Int J Dev Biol*(42), 117-125.
- Wilcock, A. C., Swedlow, J. R. y Storey, K. G. (2007). Mitotic spindle orientation distinguishes stem cell and terminal modes of neuron production in the early spinal cord. *Development*(134), 1934-1954.
- Young, P. E., Richman, A. M., Ketchum, A. S. y Kiehart, D. P. (1993). Morphogenesis in *Drosophila* requires nonmuscle myosin heavy chain function. *Genes Dev*, 29-41.
- Zhu, H. X., Thorpe, S. M. y Windle, A. H. (2001). The geometrical properties of irregular two-dimensional Voronoi tessellations. *Philosophical Magazine A*, 81, 2765-2783.

Anexo

Anexo I: Funciones polinómicas de ajuste a la vía CVT

Se presentan las 20 funciones polinómicas (2º, 3º y 4º Orden) tomadas para cada una de las realizaciones de la vía CVT. "X" representa porcentaje de hexágonos de los diagramas de Voronoi en la vía CVT. "Y" representa el porcentaje de polígonos con "N" número de vecinos.

Tipo	Funciones
2º Orden	$Y=P_1 \cdot X^2 + P_2 \cdot X + P_3$
3º Orden	$Y=P_1 \cdot X^3 + P_2 \cdot X^2 + P_3 \cdot X + P_4$
4º Orden	$Y=P_1 \cdot X^4 + P_2 \cdot X^3 + P_3 \cdot X^2 + P_4 \cdot X + P_5$

Polinomios para el cálculo de porcentaje de células con 4 vecinos

Realización	Tipo	P ₁	P ₂	P ₃	P ₄	P ₅
1	3º Grado	-0.00014954	0.031233	-2.1764	50.562	
2	3º Grado	-9.35E-05	0.021927	-1.6544	40.525	
3	3º Grado	-0.00041605	0.070696	-4.0689	80.571	
4	4º Grado	-3.13E-06	0.0006844	-0.04665	0.86149	8.9055
5	3º Grado	-0.00045815	0.080362	-4.7253	93.949	
6	4º Grado	1.10E-05	-0.0027117	0.25304	-10.59	168.04
7	3º Grado	-0.0003975	0.071847	-4.2931	85.028	
8	3º Grado	-2.90E-04	0.055845	-3.6107	78.527	
9	4º Grado	1.17E-05	-0.0020571	0.1432	-5.0905	82.591
10	4º Grado	-1.08E-05	0.0021814	-0.15	3.7417	-15.672
11	3º Grado	0.00021866	-0.025501	0.58902	9.1457	
12	3º Grado	-0.00043419	0.080461	-5.0163	105.58	
13	3º Grado	-7.24E-05	0.017773	-1.4284	37.703	
14	3º Grado	-0.00017199	0.035173	-2.3959	54.427	
15	3º Grado	-0.00055071	0.098633	-5.8968	118.33	
16	3º Grado	-0.00037466	0.068052	-4.1309	84.327	
17	3º Grado	-0.00017099	0.034475	-2.311	51.477	
18	4º Grado	3.51E-06	-0.0011902	0.13408	-6.3071	107.7
19	3º Grado	-0.0002438	0.046912	-3.0463	66.913	
20	3º Grado	-0.00035455	0.069997	-4.5567	97.835	

Polinomios para el cálculo de porcentaje de células con 5 vecinos

Realización	Tipo	P ₁	P ₂	P ₃	P ₄	P ₅
1	3º Grado	0.00054901	-0.087061	4.1454	-34.603	
2	3º Grado	-3.17E-05	-0.0021989	0.159	25.415	
3	4º Grado	-5.58E-05	0.011928	-0.93726	31.528	-351.86
4	3º Grado	8.51E-05	-0.020201	1.0116	14.11	
5	4º Grado	-3.58E-05	0.0075835	-0.5942	20.002	-215.33
6	3º Grado	6.00E-04	-0.10608	5.6414	-65.739	
7	3º Grado	0.00043073	-0.076005	3.8832	-32.178	
8	3º Grado	0.00036465	-0.066342	3.5036	-29.69	
9	4º Grado	-5.8134E-05	0.010288	-0.67136	19.013	-171.62
10	3º Grado	0.00025487	-0.045716	2.2854	-7.7752	
11	4º Grado	-1.1124E-06	3.0206E-05	0.010456	-1.0227	50.953
12	3º Grado	0.00072092	-0.12844	7.0359	-95.744	
13	3º Grado	3.17E-05	-0.0092493	0.3621	24.414	
14	3º Grado	0.00035131	-0.059302	2.8154	-12.143	
15	3º Grado	0.0004548	-0.080444	4.2487	-43.761	
16	3º Grado	0.00036038	-0.065304	3.3912	-26.41	
17	3º Grado	0.00042556	-0.07311	3.643	-27.897	
18	3º Grado	0.00086035	-0.13776	6.7407	-74.19	
19	3º Grado	0.00034158	-0.065171	3.5325	-31.115	
20	3º Grado	0.00044927	-0.083113	4.5366	-48.541	

Polinomios para el cálculo de porcentaje de células con 7 vecinos

Realización	Tipo	P ₁	P ₂	P ₃	P ₄	P ₅
1	4º Grado	-5.70E-05	0.01005	-0.65369	18.556	-170.49
2	3º Grado	0.00037348	-0.063969	3.256	-27.419	
3	3º Grado	-4.35E-05	-0.0072688	1.0324	-6.5613	
4	3º Grado	7.51E-05	-0.018829	1.1538	0.82609	
5	3º Grado	0.00059276	-0.10718	6.0234	-83.873	
6	3º Grado	0.00014137	-0.028215	1.6574	-10.548	
7	3º Grado	0.00027492	-0.052616	2.9461	-28.329	
8	3º Grado	7.53E-05	-0.017937	1.0524	2.9719	
9	2º Grado	-0.015353	1.3551	-6.7754		
10	3º Grado	0.00010761	-0.032217	2.3295	-27.763	
11	3º Grado	-6.42E-06	-0.017396	1.8223	-25.005	
12	3º Grado	-0.00037805	0.052897	-2.455	57.801	
13	3º Grado	8.81E-05	-0.024188	1.555	-6.0256	
14	3º Grado	-0.00012543	0.0023775	0.70373	-3.0917	
15	3º Grado	0.00071785	-0.12808	7.178	-106.46	
16	3º Grado	2.81E-05	-0.013278	0.98339	1.5617	
17	4º Grado	-3.30E-06	0.00053524	-0.038999	1.4478	0.19567
18	2º Grado	-0.0096436	0.83649	2.9509		
19	3º Grado	-0.00026554	0.038044	-1.8657	50.91	
20	3º Grado	0.0003466	-0.070224	4.3044	-62.499	

Polinomios para el cálculo de porcentaje de células con 8 vecinos

Realización	Tipo	P ₁	P ₂	P ₃	P ₄	P ₅
1	2º Grado	0.004176	-0.6673	25.397		
2	3º Grado	-2.25E-05	0.0070683	-0.75353	25.938	
3	2º Grado	0.0036784	-0.57955	22.305		
4	3º Grado	7.06E-05	-0.007548	-0.023478	14.264	
5	4º Grado	-0.00037732	0.063617	-3.6393	71.88	
6	3º Grado	0.00010521	-0.013633	0.29316	10.229	
7	3º Grado	-0.00017771	0.033754	-2.2055	50.318	
8	3º Grado	3.25E-05	-0.0039073	-0.052177	11.551	
9	2º Grado	0.00029378	-0.038459	1.3859	-7.5316	
10	3º Grado	-1.05E-05	0.0059241	-0.72584	25.417	
11	4º Grado	9.70E-05	-0.0079196	-0.23342	22.03	
12	3º Grado	0.00042921	-0.064975	2.9624	-36.005	
13	2º Grado	0.0037696	-0.57138	21.75		
14	2º Grado	0.0040809	-0.62388	23.606		
15	2º Grado	0.007719	-1.0529	35.953		
16	2º Grado	0.0018838	-0.36171	16.466		
17	3º Grado	0.00020348	-0.02642	0.80659	3.5065	
18	2º Grado	0.0017059	-0.39692	18.64		
19	2º Grado	0.0014519	-0.36011	18.045		
20	3º Grado	0.00011776	-0.010818	-0.086515	19.081	

Anexo II: Software suplementario

Software desarrollado en Matlab.

❖ Generación de semillas aleatorias

Función que obtiene k coordenadas (x,y) para señalar un punto en el plano con resolución $imax1 \times imax2$. Variable “length” se usa para indicar que distancia debe mantener la semilla generada con las semillas previas generadas.

```
function [m] = RandSinRepeticion(imin,imax1,imax2,K,length)
% Resuelve K valores aleatorios entre imin and imax
if ((imax1-(imin-1))*(imax2-(imin-1)) < K)
    fprintf(' Error: Excede el rango\n');
    m = NaN;
    return
end

m(1,1) = randi([imin,imax1],1);
m(1,2) = randi([imin,imax2],1);
n=1;
while (n<=K-1)
    a = randi([imin,imax1],1);
    b = randi([imin,imax2],1);
    dato=[a,b];
    for i=1:size(m,1)
        distance=sqrt((m(i,1)-a)^2+(m(i,2)-b)^2);
        if distance<=length
            ind(i)=1;
        else
            ind(i)=0;
        end
    end
    if sum(ind)==0
        dato=[a,b];
        m = [m; dato];
        n=n+1;
    end
end

m = m(1:end,:);
```

❖ Diagrama de Voronoi

Función que obtiene diagrama de Voronoi con una resolución “Resolution” empleando la variable seed como información para generar las células de Voronoi.

```
function [L_Voronoi] = Voronoi_Diagram(Resolution,seed)

%% Generación Diagrama de Voronoi %%
% Entrada : Resolution: Dimensiones de la imagen (en pixels).
%          seed: {número de célula, posición en la fila , posición en la columna}
% Salida : L_Voronoi: Diagrama de Voronoi

img = zeros(Resolution, Resolution);
for x = 1: size(seed,1)
    img(seed (x,2), seed (x,3)) = 1;
end
D = bwdist(img);
DL = watershed(D);
bgm = DL == 0;
Voronoi = bgm;
L_Voronoi = bwlabel(1-Voronoi,8); % Células numeradas
```

❖ Creación vía CVT mediante diagramas de Voronoi

Función que ejecuta “RandSinRepeticion” para generar “N_cells” semillas aleatorias distribuidas en un plano con resolución “RSxRS”. Se generan “NI” realizaciones para cada diagrama de Voronoi. Se ejecuta proceso iterativo en el que se ejecuta la función “Voronoi_Diagram”, se nombran las células y se ejecuta el algoritmo de Lloyd hasta obtener “ND” diagramas de Voronoi.

```
function DIAGRAMS_CVT_PATH(NI,RS,N_cells,ND)

%% Variables
%% NI = 20;    % Número de realizaciones
%% RS = 1024; % Resolución
%% N_cells = 500; % Número de células
%% ND = 200;  % Número de diagramas de Voronoi

%% Generación de imagen
for image = 1: NI
    % Generación de semillas para la realización "image"
    overlap = sqrt(2); % Evita colocar 2 semillas juntas
    seed_initial = RandSinRepeticion(1,RS,RS,N_cells,solape);
    seed_total{image,1} = seed_initial;
    seed_old = [];
    seed_old(:,2:3) = seed_initial;

    for diagram = 1 : ND
        %% Generación de diagram de Voronoi %%
        L_Voronoi = Voronoi_Diagram(RS, seed_old);
        n_max_cell = max(max(L_Voronoi)); % Número de células en la imagen

        %% Numerar células en la imagen en caso de diagrama distinto de 1 %%
        if diagram ~= 1
            aux = L_Voronoi;
            for x = 1: n_max_cell
                aux(L_Voronoi==L_Voronoi(seed_old(x,2),seed_old(x,3)))=seed_old(x,1);
            end
            L_Voronoi = aux;
        end

        %% Nuevas semillas %%
        center = regionprops(L_Voronoi,'Centroid');
        seed_new = cat(1, center.Centroid);
        seed_new = round(seed_new);
        seed_new = fliplr(seed_new); % [row,col]
        seed_new = [(1:n_max_cell)', seed_new];

        %% Ordenar semillas antiguas en diagram 1 (seed initial) %%
        n_cell = [];
        aux_seed = seed_old;
        if bucle == 1
            for x = 1 : size(seed_old,1)
                n_cell(x,1) = L_Voronoi(seed_old(x,2),seed_old(x,3));
                aux_seed(x,1) = n_cell(x,1);
            end
            seed_old = sortrows(aux_seed,1);
        end

        %% Salvar datos %%
        name=strcat('Data_image_',num2str(image),'_Diagram_',num2str(diagram),'.mat');
        save(name,'L_Voronoi','seed_old','seed_new')

        %% Asignar semillas del próximo diagrama de CVT path (Algoritmo de Lloyd)
        seed_old = seed_new;
    end
end
```

❖ Extracción de características en tejidos naturales y diagramas de Voronoi

Función que extrae la información de área y de vecinos correspondiente a las células válidas.

```
function [Area,N_Neighbour_cell,valid_cells] = feature_extraction(Img)

%% Generación diagramas de voronoi %%
% Entrada : Img: Tejido con células segmentadas y correctamente nombradas
n_max_cell = max(max(Img)); % Cantidad de células en la imagen
[H,W] = size(Img);
%% Detección de área %%
Area = regionprops(Img,'Area');
Area = cat(1, Area.Area);
%% Detección de vecinos %%
radio=4;
for cell=1 : n_max_cell
    BW2 = bwperim(Img == cell);
    [pi,pj] = find(BW2 == 1);
    se = strel('disk',radio);
    BW2_dilate = imdilate(BW2,se);
    pixels_Neigh = find(BW2_dilate == 1);
    Neighbour = unique(Img(pixels_Neigh));
    Neighbour = Neighbour(find(Neighbour ~= 0 & Neighbour ~= cell));
    Neighbour_cell{cell} = Neighbour;
    N_Neighbour_cell(cell) = length(Neighbour_cell{cell});
end
%% Células Válidas %%
% Detección de células en el borde de la imagen
Img_det_edge = bwlabeled(Img,8);
Img_det_edge(1,1:W) = 1;
Img_det_edge(H,1:W) = 1;
Img_det_edge(1:H,1) = 1;
Img_det_edge(1:H,W) = 1;
Img_det_edge = bwlabeled(Img_det_edge,8);
cells_no_edge=[];
cells_edge=[];
for cell = 1: n_max_cell
    if unique(Img_det_edge(Img == cell)) ~= 1
        cells_no_edge = [cells_no_edge,cell];
    else
        cells_edge = [cells_edge,cell];
    end
end
% Células no válidas (células del borde y sus vecinos)
valid_cells = cell_no_edge;
not_valid_cells = cells_edge;
for cell = 1 : length(cells_edge)
    Neigh_edge_cell = Neighbour_cell{cells_edge(cell)};
    flag = 0;
    for j = 1 : length(Neigh_edge_cell)
        valid_cells = valid_cells (find(Neigh_edge_cell(j)~= valid_cells));
        not_valid_cells = [not_valid_cells, Neigh_edge_cell (j)];
    end
end
not_valid_cells_previous = unique(sort(not_valid_cells));
valid_cells_previous = unique(sort(valid_cells));
% Eliminar células válidas aisladas
valid_cells = valid_cells_previous;
for j = 1 : length(valid_cells_previous)
    Neigh_cell_j = Neighbour_cell{valid_cells_previous(j)};
    no_equal = [];
    for x=1:length(Neigh_cell_j)
        no_equal(x) = isempty(find(Neigh_cell_j(x) == not_valid_cells_previous));
    end
    if sum(no_equal) == 0
        valid_cells = valid_cells(find(valid_cells ~= valid_cells_previous(j)));
        not_valid_cells = sort([valid_cells_previous(j),not_valid_cells_previous]);
        flag = 1;
    end
end
if flag == 0
    valid_cells = valid_cells_previous;
    not_valid_cells = not_valid_cells_previous;
end
```

❖ Extracción de características en diagramas de la vía CVT

En los diagramas de Voronoi de la vía CVT las células válidas precisan ser las mismas a lo largo de las “ND” diagramas. El programa extrae esta información para adquirir las células válidas en las “NI” realizaciones.

```
function feature_extraction_cvt_path (NI,ND,Max_cells)
%% Variables
% NI = 20;           % Number of Images
% ND = 200;          % Number of Voronoi Diagram
% Max_cells = 500;   % Higher number of cells involved ND Voronoi Diagram

for image = 1 : NI
    Matrix_valid_cells=zeros(ND, Max_cells);
    for diagram = 1 : ND
        name=strcat('Data_image_',num2str(image),'_Diagram_',num2str(diagram),'.mat');
        load(name)
        [Area,N_Neighbour_cell,valid_cells] = feature_extraction(L_Voronoi);
        Matrix_valid_cells(diagram, valid_cells)=1;
        save(name,'L_Voronoi','seed_old','seed_new','L_Voronoi',...
            'Area','N_Neighbour_cell')
    end
    valid_cells_path = sum(Matrix_valid_cells,1);
    valid_cells_path = find(valid_cells_path == ND);
    save(strcat('valid_cells_path_image_',num2str(image)), 'valid_cells_path')
end
```

❖ Densidad de datos

Función que genera un gráfico de densidad con los datos [x,y] con resolución [widthxheight] y con las dimensiones [Col_{min}, Col_{max}, Fila_{min}, Fila_{max}] definidas en “limits”

```
function [ dmap ] = dataDensity( x, y, width, height, limits, fudge )
%DATADENSITY obtiene una imagen con la densidad de los datos
% x, y - dos vectores de igual longitud
% width, height - resolución de la gráfica (en pixels)
% limits - [xmin xmax ymin ymax]
% fudge - cantidad de muestras, por defecto la diagonal en pixel
%
% By Malcolm McLean
%

deltax = (limits(2) - limits(1)) / width;
deltay = (limits(4) - limits(3)) / height;
if(nargin < 6)
    fudge = sqrt (deltax^2 + deltay^2);
end
dmap = zeros(height, width);
for ii = 0: height - 1
    yi = limits(3) + ii * deltay + deltay/2;
    for jj = 0 : width - 1
        xi = limits(1) + jj * deltax + deltax/2;
        dd = 0;
        for kk = 1: length(x)
            dist2 = (x(kk) - xi)^2 + (y(kk) - yi)^2;
            dd = dd + 1 / ( dist2 + fudge);
        end
        dmap(ii+1,jj+1) = dd;
    end
end
```

❖ Gráfico de densidad de la vía CVT

Función que genera un gráfico de densidad de datos de la vía CVT del modelo continuo definido por las funciones polinómicas (Anexo I) almacenadas en la variable “parameters functions”. Se obtiene un grafico por cada tipo de polígono analizado (4, 5, 7 y 8).

```
load('parameters_functions','functions','parameters')
% Function is a variable of class "cell" each cell contain the classes of functions to
% evaluate percentage of (4,5,7 or 8 polygons) in a vector of length NI. Each
% cell from the vector have the value 0 , 1 or 2.
% 0: cvt Path of image "image" get with a quadratic polynomial function
% 1: cvt Path of image "image" get with a cubic polynomial function
% 2: cvt Path of image "image" get with a 4° degree polynomial function
% Parameter is a variable of class "cell" each cell contain a matrix of size [NIx5].
% Each row from the matrix have the parameters to perform function defined in
% function.
NI = 20;
Percentage = [4,5,7,8];
Xmin = 20;
Xmax = 60;
Ymin = -5;
Ymax = 50;
Xadd = Xmax-Xmin;
for count = 1 : length(Percentage)
    x1 = [];
    y1 = [];
    for img = 1 : NI
        Axis_x = rand(500,1).*Xadd+Xmin; %random number between xmin and xmax
        if functions{Percentage(count),1}(img) == 0 %Quadratic
            P = parameters{count,1}(img,1:3);
            x = Axis_x;
            x1 = [x1;x];
            y = P(1,1).*x.^2 + P(1,2).*x + P(1,3);
            y1 = [y1;y];
        elseif functions{Percentage(count),1}(img) == 1 %Cubic
            P = parameters{count,1}(img,1:4);
            x = Axis_x;
            x1 = [x1;x];
            y = P(1,1).*x.^3 + P(1,2).*x.^2 + P(1,3).*x + P(1,4);
            y1 = [y1;y];
        elseif functions{Percentage(count),1}(img) == 2 %4° degree
            P = parameters{count,1}(img,1:5);
            x = Axis_x;
            x1 = [x1;x];
            y = P(1,1).*x.^4 + P(1,2).*x.^3 + P(1,3).*x.^2 + P(1,4).*x + P(1,5);
            y1 = [y1;y];
        end
    end
end
%% Density Plot
levels = 256;
limits = [Xmin Xmax Ymin Ymax];
map = dataDensity(x1, y1, 256, 256, limits );
map = map - min(min(map));
map = floor(map ./ max(max(map)) * (levels-1));
map = flipud(map);
map = 255-map;
f = figure();
%% Coordinate definition
image(map);
colormap(gray(levels));
vxlabel = [Xmin : 5 : Xmax];
vxpos = ((levels-1)/Xadd).*vxlabel-(((levels-1)/Xadd)*Xmin)-1);
vylabel = [Ymax : -5 : 0];
vypos = -((levels-1)/(Ymax-Ymin)).*vylabel+(((levels-1)/(Ymax-Ymin))*Ymax)+1);
set(gca,'XTick' vxpos);
set(gca,'XTickLabel',vxlabel);
set(gca,'YTick',vypos);
set(gca,'YTickLabel',vylabel);
end
```

Anexo III: Artículo publicado correspondiente al capítulo I de resultados

Título: Topological Progression in proliferating epithelia is driven by a unique variation in polygon distribution

Revista: Plos One

Autores:

- Daniel Sánchez-Gutiérrez
- Aurora Sáez
- Alberto Pascual
- Luis M. Escudero

Año: 2013

Topological Progression in Proliferating Epithelia Is Driven by a Unique Variation in Polygon Distribution

Daniel Sánchez-Gutiérrez¹, Aurora Sáez², Alberto Pascual^{1,3}, Luis M. Escudero^{1,3*}

1 Instituto de Biomedicina de Sevilla, Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville, Spain, **2** Escuela Técnica Superior Ingeniería, Universidad de Sevilla, Seville, Spain, **3** Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain

Abstract

Morphogenesis is consequence of lots of small coordinated variations that occur during development. In proliferating stages, tissue growth is coupled to changes in shape and organization. A number of studies have analyzed the topological properties of proliferating epithelia using the *Drosophila* wing disc as a model. These works are based in the existence of a fixed distribution of these epithelial cells according to their number of sides. Cell division, cell rearrangements or a combination of both mechanisms have been proposed to be responsible for this polygonal assembling. Here, we have used different system biology methods to compare images from two close proliferative stages that present high morphological similarity. This approach enables us to search for traces of epithelial organization. First, we show that geometrical and network characteristics of individual cells are mainly dependent on their number of sides. Second, we find a significant divergence between the distribution of polygons in epithelia from mid-third instar larva versus early prepupa. We show that this alteration propagates into changes in epithelial organization. Remarkably, only the variation in polygon distribution driven by morphogenesis leads to progression in epithelial organization. In addition, we identify the relevant features that characterize these rearrangements. Our results reveal signs of epithelial homogenization during the growing phase, before the planar cell polarity pathway leads to the hexagonal packing of the epithelium during pupal stages.

Citation: Sánchez-Gutiérrez D, Sáez A, Pascual A, Escudero LM (2013) Topological Progression in Proliferating Epithelia Is Driven by a Unique Variation in Polygon Distribution. PLoS ONE 8(11): e79227. doi:10.1371/journal.pone.0079227

Editor: Eduardo Moreno, University of Bern, Switzerland

Received: July 4, 2013; **Accepted:** September 27, 2013; **Published:** November 5, 2013

Copyright: © 2013 Sánchez-Gutiérrez et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: LME is supported by the Miguel Servet (Instituto Carlos III) program that also funded the work. LME and DSG are funded by the Spanish Ministry of Science (BFU2011-25734). AS is funded by the Consejería de Innovación, Ciencia y Empresa of the Junta de Andalucía. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lmesudero-ibis@us.es

Introduction

Epithelia are one of the fundamental units of animal development. These tissues undergo cell shape changes and reorganizations within the epithelial plane that sculpt the final organism [1,2]. Remodeling is a finely controlled process that involves different types of rearrangements. Gradually, small reorganizations vary the topology of the whole epithelia. Despite being a highly dynamic process, it also has to be ordered and fairly reproducible enabling the correct formation of mature organs with a cellular assembly suited to their specialised functions. Combination of genetics and imaging techniques has enabled profound advances in the understanding of fast and dramatic morphogenetic events in *Drosophila*. Clear examples of them are the rearrangements induced during convergent extension or ommatidia rotation, the tissue remodeling dependent on cell apical constriction or the zippering of the embryo during the process of dorsal closure [3,4,5,6,7,8,9,10,11,12,13].

The wing primordium has been particularly well studied in terms of the genetic inputs that drive its development. This great understanding has made its monolayer epithelium the perfect target to analyze the development of epithelial topology. During the four days of larval development, wing imaginal disc grows from 20 to approximately 50000 cells [14,15]. Later in development, at pupa stages, cell divisions stop and this epithelium became an almost perfect hexagonal lattice that will originate the

adult wing. Interestingly, through this period of intense proliferation the topology of the epithelium seems heterogeneous with not apparent order or governing organizational rules. Several groups have tried very different approaches to understand the mechanisms of cells packing during the development of the wing disc. Remarkably, all of them are based directly or indirectly in the number of sides of the cells and detected similar values of the polygon distribution in proliferating wing discs [1,2,16,17,18,19,20,21,22,23,24]. Different approaches have been used so far to understand the basis of the emergence of epithelial topology and cellular packing. They goes from genetic analysis of morphological changes [19], the analysis of the interplay between of proliferation, mitotic cleavage and topology [17,23,25], computational modelling, [21,22] or the study of the biophysical properties of the cells proposing a vertex model [16,24]. These studies imply the existence of a proliferative phase not involved in organization and a second non-proliferative phase (after pupa formation) where final order is acquired.

Here we try to address if the proliferative phase can already contribute to epithelial organization, and if that is the case, to define what are the main organizational clues that arise during this early developmental time point. In order to do that we have performed for the first time a detailed comparison between two developmental points during the proliferative phase: mid-third instar larva and early prepupa (separated only by 24 hours of development). We demonstrate the existence of differences

between these two close stages of development using several systems biology methods (including our new network-based image analysis approach, [26,27,28] and a high number of samples (31 samples and a total of 15951 cells). The network characteristics that discriminate between these two stages provide also new biological information about this developmental process. Using these methods, we are able to detect the emergence of homogeneity and regularity before the end of the larval proliferative stage. This represents the first hints of hexagonal packing that will occur on the pupal wing disc.

Results

Computerized analysis shows that cellular characteristics do not vary between larval and prepupal samples.

We have used *Drosophila* wing epithelium as a model to understand the mechanisms that rule epithelial organization. Our dataset consisted in 15 samples from middle third instar wing discs (dWL), and 16 samples from the same region of early prepupa wing discs (dWP). These two proliferative stages are separated by 24 h of development. We have used segmented images from our previous study [26] in order to identify the objects (cells) that compose them. This enables the obtaining of geometric and topological characteristics from every cell. During the passage from mid third instar to prepupa there are only small changes at the level of the apical surface of the wing disc cells. Therefore these images appear very similar by simple visual inspection (**Fig. 1A**).

In order to identify subtle differences between dWL and dWP cells, we have tested a series of 14 characteristics that could reflect differences at the cellular level between our two types of images (**Table 1**). 1 to 5 were geometric features of the cells, while 6 to 14 corresponded to topological features capturing different aspects of the relation of each cell with its neighbours. Topological features were extracted constructing a network of cells contacts; with the centroid of the cells being nodes and adjacent cells being connected [26] and **Methods**). Once that we extracted the values for these 14 characteristics, we performed a correlation analysis using all the epithelial cells (dWL = 9070 cells, dWP = 6881 cells). As a result, we obtained a matrix showing the degree of similarity between every pair of cells. This dataset was interpreted using a network representation. Cells with a correlation coefficient above a determined threshold were connected building a “network of correlating cells” (**Fig. 1B-C**).

Our hypothesis was that if these 14 features were able to differentially characterize cells from one particular stage we should find clusters of dWL cells and clusters of dWP cells. This was not the case. The largest correlation graphs always contained a mix of dWL and dWP cells (independently what threshold was used) (**Fig. 1B**). Our cell correlation experiments suggest that dWL and dWP cells do not present characteristics that allow their separation depending on the developmental stage. However, despite the fact that “number of cell sides” was not one of the characteristics used for the correlation, the combination of dWL and dWP cells in the network presented a clear preference in the distribution that depended on the type of polygon (**Fig. 1C**). Therefore, the “number of cell sides” bias the values for the 14 characteristics analyzed giving a certain local constrain to the tissue. These results encouraged us to search for differences at the next level, the polygon distribution.

Drosophila wing epithelia change the polygon distribution during larval development.

Previous studies have shown that the apparition of a determined polygon distribution in the wing disc of *Drosophila* (around 3%

tetragons, 28% pentagons, 46% hexagons, 20% heptagons) is an inherent property of the proliferating epithelia and is present in other metazoan [17]. We calculated the percentage of cells with different number of sides in the two developmental stages analyzed (dWL and dWP; **Fig. 2A**). The average values were in the same range to the previously published [17,22]. However, we found a small but significant difference between dWL and dWP polygon distribution (MANOVA test, $p = 0.013$, **Fig. 2A**). For example, dWL presented a lower number of hexagons and a higher number of pentagons than dWP. Hence, we conclude that developmental factors drive a polygon distribution variation during this proliferative period. One of these developmental factors could be the reduction of proliferation rate that occurs at the end of the larval stage [14,15,29].

Fast developmental changes in epithelial organization are revealed by network characteristics of groups of cells.

An arising question was if the change in polygon distribution might lead to a progression in epithelial organization that could be detected using independent approaches. In a recent work, we have shown that differences in epithelial organization can be captured using “network” characteristics stemming from images with a large number of cells [26]. These topological characteristics are extracted from the network of cellular contacts described before where every cell of the epithelium is considered a node that is connected with the adjacent cells [26]. Aiming to improve our previous method, we have designed a new set of 40 characteristics (**Table 1**). 32 of these characteristics comprised concepts from graph theory and complex networks [27]. After this, we used Principal Component Analysis (PCA) to analyze the results. PCA is an unbiased mathematical algorithm that lowers data dimensionality, such that it projects the data on orthogonal axes maximizing their dispersion. In our case the data are the feature vectors which represents the images under study. Thus, images can be plotted in a bi-dimensional space representing the two first principal component. An analysis of the components plots can then be used to identify similarities and differences between images. In this way, very different images will distribute far from each other, while similar images will cluster together. This graph allows the quantification of the differences between the groups formed for each type of data (dWL and dWP in our case). We have designed a feature selection step to obtain only the most relevant characteristics to discriminate between two types of images ([28] and **Methods**). This feature selection step performed different combinations of the 40 characteristics and tested which resulted in a best separation in the PCA graph. To calculate the degree of separation we used a “PCA descriptor” that gave us a numeric value for each PCA graph resulting from different combinations of characteristics (in our case from 0.1 to 1.8 approximately, see **Methods**). Therefore, the aim is to find the combination of characteristics that maximize this “PCA descriptor”. In the case of dWL-dWP comparison the program selected one geometric and four “network” features: “Average major Axis”, “Average Relation Neighbours Convex Hull”, “Average Betweenness Centrality”, “Standard Deviation Strength” and “Efficiency” (see **Discussion** and **Text S1**). Using these characteristics the PCA graphs showed dWL and dWP images separated with only a small region of overlapping (**Fig. 2B**). The corresponding value for the PCA descriptor was 1.527.

This result suggested that our approach was able to identify topological differences between these two developmental stages. However, one possibility was that our feature selection step was able to force the separation between both sets (resting biological relevance to the experiment). To discard this possibility, we

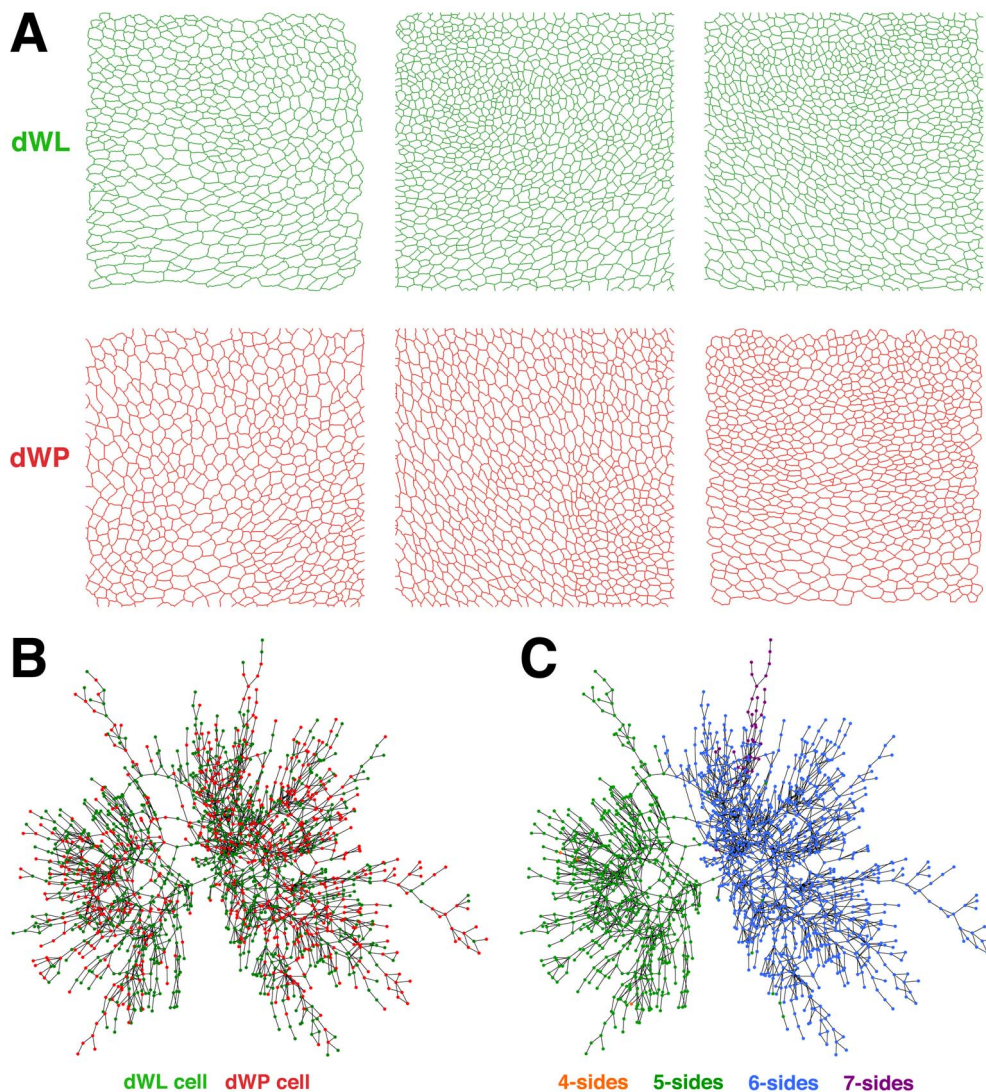


Figure 1. Epithelial images and cell correlation. **A)** Segmented image of three dWL (green) and dWP (red) images showing the extreme similarity between them. **B)** Visualization of the results of the correlation between dWL (green) and dWP (red) cells. The network contains cells from both types of images and each cell is represented by a node. Two nodes are linked if they present a similarity bigger than a certain threshold. The network shown in the panel is the one with a higher number of nodes (1729 cells) using a threshold of 0.9975. **C)** Representation of the same network of panel **B**, showing the distribution of sides of each cell. Orange, green, blue and purple mark 4, 5, 6, and 7 sided cells respectively. The image shows the high tendency of cells with the same number of sides to be linked.
doi:10.1371/journal.pone.0079227.g001

designed an assay to generate random combinations of our images, perform the feature selection step, and calculate the PCA descriptor value (see **Methods**). The results for 4,000 loops showed that the value obtained for dWL-dWP comparison was higher than 99.9% of the random combinations (**Fig. S1**). This latter result supported the existence of real (and detectable) differences between dWL and dWP images.

Relation between the change in epithelial polygon distribution and the emergence of a different organization

So far, we have shown that dWL and dWP images present a different polygon distribution and a different epithelial organization. It might be possible that the latter was simply a consequence of the significant variation in polygon distribution. Under the light of our experiments, this is not the case. First, we obtained the

polygon distribution from the 4,000 random combinations used previously and performed a MANOVA test for each combination. We found that only in 32 cases (0.08%) the achieved p -value was lower than the obtained in the dWL-dWP comparison (**Fig. S2**). Second, we plot the p -values of the MANOVA test and the PCA descriptor results (**Fig. 2C**). The graph showed that there was not a general correlation between these two parameters: We found that the 4 cases with a better value for the PCA descriptor did not present differences in the polygon distribution. On the other hand, the value of the PCA descriptor for the arbitrary groups with bigger differences in their polygons distributions (MANOVA, p -value < 0.05) were always smaller than 1. In summary, both features seem to be independent from each other.

Consequently, only in the case of the comparison dWL-dWP a high value of the PCA descriptor corresponded to a low p -value of the MANOVA test (**Fig. 2C**) suggesting the existence of

Table 1. List of characteristics analyzed in this study.

CHARACTERISTICS		
epithelial cc	Name	cell cc
1	Average Area	1
2	S. D. Area	
3	Average major Axis	2
4	Average minor Axis	3
5	Average Relation Axis	4
6	S. D. Relation Axis	
7	Average Convex Hull	5
8	S. D. Convex Hull	
9	Average Neighbours	
10	S. D. Neighbours	
11	Average Relation Neighbours Area	6
12	S. D. Relation Neighbours Area	
13	Average Relation Neighbours major axis	7
14	S. D. Relation Neighbours major axis	
15	Average Relation Neighbours minor axis	8
16	S. D. Relation Neighbours minor axis	
17	Average Relation Neighbours relation axis	9
18	S. D. Relation Neighbours relation axis	
19	Average Relation Neighbours convex hull	10
20	S. D. Relation Neighbours convex hull	
21	Average Strengths	11
22	S. D. Strengths	
23	Average Clustering Coefficient	12
24	S. D. Clustering Coefficient	
25	Average Eccentricity	13
26	S. D. Eccentricity	
27	Average Betweenness Centrality	14
28	S. D. Betweenness Centrality	
29	Average Shortest Paths lengths	
30	S. D. Shortest Paths Lengths	
31	Radius	
32	Diameter	
33	Efficiency	
34	Pearson correlation	
35	Algebraic connectivity	
36	S _{metric}	
37	Assortativity	
38	Density	
39	Transitivity	
40	Modularity	

Table shows names of the 40 characteristics analyzed in the feature selection step by PCA descriptor (a description is included in the **Text S1**). The 40 characteristics can be classified into three types: geometrically related to the size and shape of cells (1–8), network characteristics of the cells (9–28) and network characteristics of the image (29–40). The network features capture information about the organization of the cells. The grey background marks the 14 characteristics used in the cell correlation assay (numeration is on the right side).

doi:10.1371/journal.pone.0079227.t001

developmental mechanisms that links polygon distribution changes with detectable organizational transformations.

Discussion

The analysis of small global changes during development is key to understand the mechanisms that couple growth and morphogenesis. We have used the well-studied *Drosophila* wing disc epithelia as a test-tube, comparing geometric and topological properties of mid-third instar larva and early prepupa stages, two time-points without recognizable organizational patterns (**Fig. 1A**). The acquisition of the hexagonal packing leading to the final wing organization has been described as dependent on planar cell polarity genes during pupa stages [19]. Therefore, the last hours of larval development has been classically seen as an actively growing stage lacking coordinated rearrangements beyond the maintaining of a specific polygonal distribution [17]. Our results suggest the existence of small but significant topological differences between mid-third instar larva and prepupa stages, marking the onset of the reorganization at the proliferative phase.

We started searching differences at the level of the individual cells. Using a correlation approach we aimed to compare the similarities between all the cells that integrated our two groups of images. The similarity analysis produced a clear result: nodes (cells) were not grouped depending of the original developmental stage (**Fig. 1B**). On the contrary, the mapping was completely related to the number of sides of the cells: Six side cells were preferentially linked to other hexagons, pentagons were clustered together and also groups of heptagons were favoured (**Fig. 1C**). This experiment raises a clear conclusion; the values of the 14 characteristics used in the correlation (**Table 1**) were strongly dependent on the number of sides of the cells (although “number of sides” was not one of the 14 characteristics used), and independent of the developmental stage. This dependence affected not only geometric, but also topological characteristics of the individual cells. On the other hand, we found that the polygon distributions of dWL and dWP were significantly different. Putting together these two findings we hypothesize that if the number of sides of a cell influence its local topological characteristics, a significant alteration in the polygon distribution can be converted into a coordinated morphogenetic transformation (**Fig. 2D**). In other words, changes in the prevalence of a polygon over other could lead into the modification of the organization of the whole epithelium. However, although we favour this option, we cannot totally discard that both phenomena are interdependent and that a modification of the arrangement of the tissue would also lead to the change in polygon distribution.

The remaining question is if only a difference in polygon distribution can explain changes in epithelial organization. We have tested 4000 combinations of arbitrary groups of images to conclude that polygon distribution and epithelial organization are, therefore, independent features in the random combinations. Consequently, only our biologically meaningful combination was significant for both tests and better than other 4000 combinations (**Fig. 2C**). For this reason, we interpret that during development the change in polygon distribution that occurs between mid-third instar larva and early prepupa is highly coordinated. This enables the emergence of higher magnitude organizational changes that now can be captured with our new network based method.

These organizational changes are best described by the relevant characteristics used to differentiate dWL and dWP samples. The “Average Major Axis”, “Average Relation Neighbours Convex Hull” and “Average Betweenness Centrality” stand out as the more relevant when analyzing the order in which they appear in the

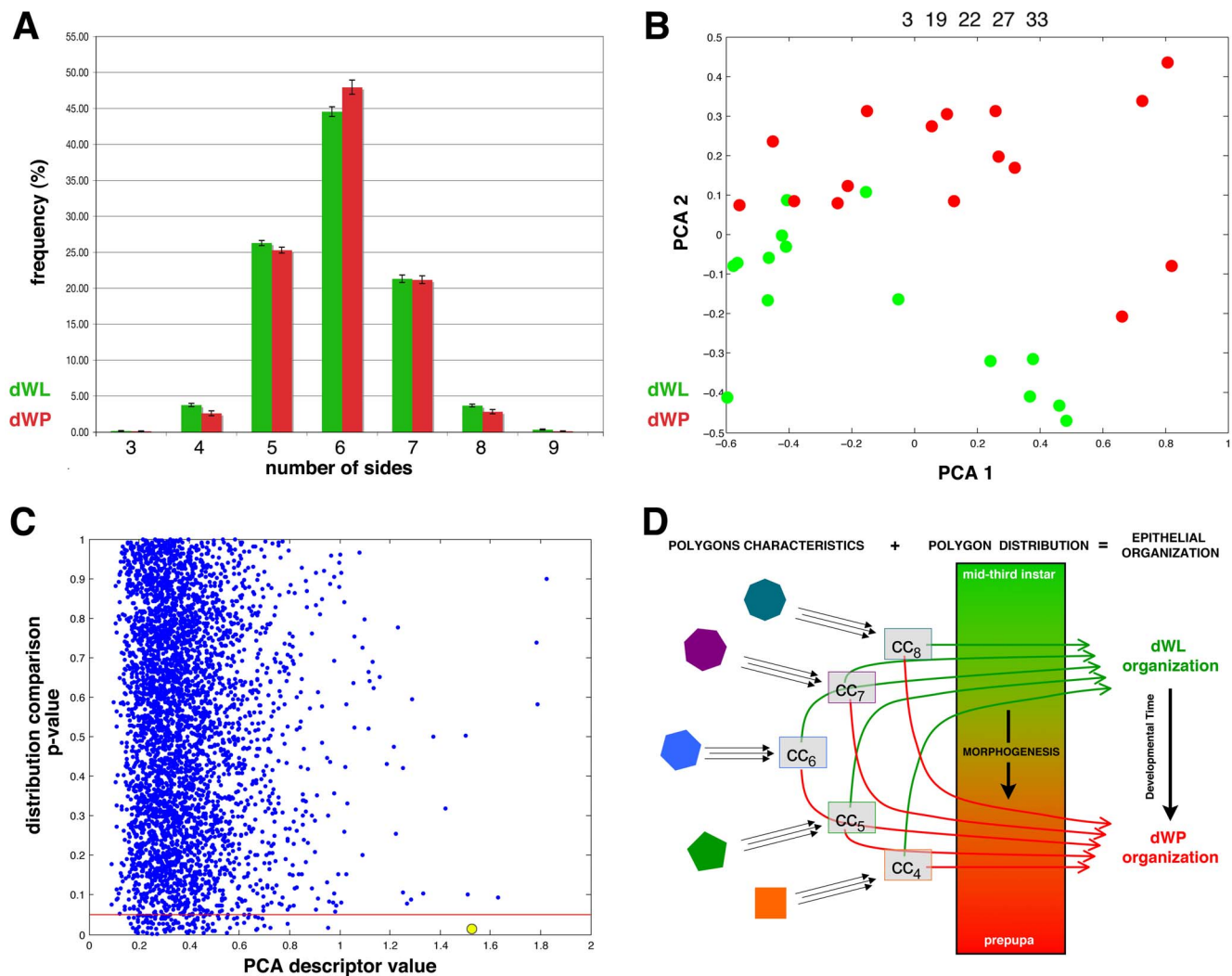


Figure 2. Epithelial organization differences between dWL and dWP. **A**) Polygon distribution of dWL (15 datapoints, green) and dWP (16 datapoints, red) images. The frequency of each type of polygons in both sets of images is represented. The error bars represent the standard error. **B**) PCA graph for the comparisons of dWL (green dots) and dWP (red dots) images using the selected characteristics (numbers 3, 19, 22, 27 and 33). **C**) Graph representing the 4000 random combinations of images (blue dots). The p -value resulting from the MANOVA test of the distribution comparison is plotted against the PCA descriptor value of the same random combination. The red line marks the $p=0.05$. The graphs show the absence of correlation between both values. The yellow circle marks the dWL-dWP combination. Only four combinations present a higher PCA descriptor value. None random combination with $p \leq 0.05$ shows a PCA descriptor greater than dWL-dWP combination. **D**) Model for control of tissue organization during the end of the proliferative larval stage. The morphogenetic signals in the wing disc drive a change in polygon distribution between mid-third instar larva and early prepupa. Our results also support the existence of two separable organizations in each of these developmental time-points. The model propose that the number of sides of the cells imposes geometric and organizational local constraints that in combination with a determined variation of the percentage of each polygon propagate in a change of the tissue epithelial organization. doi:10.1371/journal.pone.0079227.g002

feature selection step and their weights in the PCA. Interestingly, these three characteristics were also used in the correlation experiment. What is the biological meaning of them? Their definition can explain the changes during these last 24h of larval development. In the case of geometric characteristics such as average major axis, it is easy to interpret that dWP cells growth increasing the value of their major axis (the same happens with the values for minor axis" and "Average area", **Table S1**). The case of "Average Relation Neighbours Convex Hull" is more complicated. The convex hull is defined as the smallest convex set that contains the shape. However, to have a reference with respect to the cell, we computed this value as "Area of cell / convex hull area of the cell". Therefore, a value close to 1 indicates that the cell presents a convex shape with smooth and straight sides. An

increase of wiggles and irregularities in the perimeter of the cell decreases this value. Our data indicate that the outline of dWP cells is more regular: the values of "Average Convex Hull" and "Average Relation Neighbours Convex Hull" are closer to 1 (Table S1). We interpret that in dWP images the cells are more similar to their neighbours in terms of contour regularity, and this is one of the most important features to discriminate between dWL and dWP. In biological terms, this characteristic could express in some extent the effect of the mechanical stress that stretch the cells outlines. Clusters of early prepupa cells would present higher tensions induced by the morphogenetic movements that occur at this stage [29]. The third more discriminant feature is the "Average Betweenness Centrality" defined by the fraction of all shortest paths in the network that contain a given node. Nodes

with high values of betweenness centrality participate in a large number of shortest paths, they are usually called hubs [30,31]. dWL present a greater number of these hubs combined with others cells with lower value. This suggests that the tissue is more heterogeneous in terms of connectivity of their nodes, i.e. organization of their cells. On the other hand, early prepupa samples show a smaller value of “Average Betweenness Centrality” and “Std. Dev. Betweenness Centrality”. These results, translated to our images, indicate an increase of the homogeneity of this tissue with respect to dWL. The remaining two characteristics present a lower PCA weight and complement the previous three to obtain a higher PCA descriptor value. Average Efficiency and Std. Dev. Strength would be related with the increase of homogeneity and size respectively. To summarize, the relevant characteristics able to separate dWL and dWP images indicate that the enrichment of six-sided cells starts when the disc is still growing. An enlargement of the cells, a smoothened of their outlines and an increase of the homogeneity of the whole epithelium lead to the epithelial hexagonal lattice that will be formed later at pupa stage [19]. Our new methods of analysis mark the onset of the “hexagonal packing” during the proliferation stage of larval development, although it only will become obvious at the pupa stage [19].

All our results together combine in a simple model (**Fig. 2D**). We have demonstrated that the number of sides of the cells imposes local geometric and organizational constraints. Therefore changes in the percentage of each polygon can propagate altering the whole epithelial organization. Surprisingly, only the variation in polygon distribution driven by morphogenesis leads to progression in epithelial organization (**Fig. 2C**). The question remains about how the coupling of growth and organization is orchestrated. A simple hypothesis could imply the planar cell polarity genes acting before cell proliferation is arrested. Also, it could be caused by topological changes induced by the decrease of proliferation rate observed at the end of the larval period. Other possibility, related to the newly proposed integration of force-sensing and signaling pathways [32], could also explain small organizational changes. Future works should shed light on the detailed mechanisms that drive the process.

Here we have presented a new method to discriminate between very similar epithelial images that identify the relevant characteristics that allow the separation. This approach can be easily adopted for the analysis of epithelia in other systems, and could be very valuable to analyze small developmental changes visualized using *in vivo* imaging.

Materials and Methods

Image analysis and extraction of characteristics

dWL (mid third instar) and dWP (early prepupa) images were obtained at 96h and 120h of development respectively (growth at 25°C). All images analyzed in this study come from [26]. A complete description of how the images were obtained can be found in the methods section of that article. In each image a Region of Interest (ROI) was established in order to exclude cells closer to the border of the image. The features of the cells falling within the ROI were calculated. The cells outside were only used in order to provide neighbours to the cells analyzed [26]. A network of cellular contacts was created taking the centroid of each cell as a node that links to centroids of adjacent cells. To build the network, we followed the same method described in [26] with only one modification: We have used a radius $r = 4$ for the circle used to identify the neighbours of each cell.

15951 cells have been analyzed in 31 images. We defined 14 features related with geometric and topological properties of the individual cells than can be extracted from them. These values were used in the correlation assay. We also defined 40 features of the images. The values for features 1 to 28 were computed extracting the value for each cell and calculating the average and the standard deviation of all the cells inside the ROI for each image (**Table 1** and **Text S1**) [26]. The remaining twelve characteristics were extracted directly from the network formed by the nodes inside the ROI. After the extraction of the characteristics, the values were normalized to enable the comparison between different characteristics.

Correlation assay

We have performed a correlation assay to measure the similarity of the individual cells contained in our images. We have compared the values of 14 characteristics: “average (av.) area”; “av. major axis”; “av. minor axis”; “av. relation axis”; “av. convex hull”; “av. relation (rel.) of neighbours area”; “av. rel. of neighbours major axis”; “av. rel. of neighbours minor axis”; “av. rel. of neighbours relation axis”; “av. rel. of neighbours convex hull”; “av. rel. strengths”; “av. clustering coefficient”; “av. eccentricity”; “av. betweenness centrality”. The first step was the extraction of these 14 values from each one of 15951 cells, then, an exhaustive pair wise comparison was performed. Our correlation process compared cells along 14 features. If our data is defined as $cell1 = [feature\ 1, feature\ 2, \dots, feature\ 14]$ and $cell2 = [feature\ 1, feature\ 2, \dots, feature\ 14]$, the following equation provides the correlation value of these two cells:

$$Corr = \frac{\sum_{n=1}^{14} [cell1(n) - \overline{cell1}] * [cell2(n) - \overline{cell2}]}{\sigma_{cell1} * \sigma_{cell2}}$$

Where σ_{cell1} show us the standard deviation of cell1 data and $\overline{cell1}$ is the average of the normalized values of the 14 features from cell1. This measurement was done for each pair of cells. The obtained “Corr” value must be in a range between 0 and 1; the closer to one, the higher similarity between cells. We defined a threshold to get pairs of cells that are clearly similar. The threshold was applied as follow:

- If $Corr (cell1, cell2) \geq Threshold$; then $distance(cell1, cell2) = 1 \rightarrow$ Cells connected.
- If $Corr (cell1, cell2) < Threshold$; then $distance(cell1, cell2) = 0 \rightarrow$ Cells disconnected.

Visone software was used to visualize the relation between cells. This program integrated the distance data to build the “correlation network” of connected cells. Two cells were connected if their correlation value was above the threshold. To determinate an optimum threshold value we analyzed the networks created. If threshold was too high, networks would present only few related cells. In the other hand, if threshold was too low, the resulting network would be difficult to provide some information. We examined different thresholds to corroborate that our results did not depend of the chosen limit. For clarity, we selected to show a threshold that created a larger network with 1500-2000 cells (**Fig. 1 B, C**). Visone can manage different “Labels” that facilitates the analysis of the distribution of different characteristics of the cells.

Randomization

In this work we have analyzed 31 images: 15 from mid third instar larva (dWL) and 16 from prepupa stage (dWP). Our results are based in the comparison of both groups. In some of our experiments we have randomly created new pairs of groups. We separated our 31 images in two groups: mixA (8 images from dWP and 8 images from dWL) and mixB (the 8 remaining images from dWP and the 7 images from dWL). For each randomization loop, two new mixA and mixB groups were obtained and analyzed. Finally, we have compared the results obtained with the original dWL and dWP groups with the 4000 randomizations.

Polygon distribution analysis

To evaluate if the polygonal distribution of two groups of images was significantly different a Multivariate Analysis of Variance (MANOVA) was used. If $p\text{-value} < 0.05$, distributions were considered to be significantly different. We compared the polygon distribution of dWL and dWP groups and 4000 random combinations of our 31 images. We performed all these MANOVA tests using the values for cells with 4, 5, 6, 7 and 8 sides. We discarded the values for the cells with 3, 9 and 10 sides, since they were not present in all the images.

Principal Component Analysis (PCA) and PCA's Descriptor

We have used PCA [33] to analyze the differences between our two sets of images using the extracted characteristics [26]. PCA transforms the correlated data points of the feature vector into a small number of uncorrelated variables called principal components. The projection maximises the dispersion of the individual data points in an unbiased way. This allows the identification of naturally separated sets of data points (images in our case). These data points can be visualized graphically based on its position on the PCA graph when 2 principal components are represented. Once the PCA graph was obtained, we used a variant of Calinski-Harabasz descriptor to evaluate the degree of separation between 2 groups of images [34].

$$W = \sum_{i=1}^k \sum_{l=1}^{N_i} (x_l(I) - \bar{x}_i)(x_l(I) - \bar{x}_i)^T$$

$$B = \sum_{i=1}^k N_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$$

$$T = W + B = \sum_{l=1}^k (x_l(I) - \bar{x})(x_l(I) - \bar{x})^T$$

$$PCA_Descriptor = trace\left(\frac{B}{W}\right)$$

Where it is given a set $X = \{x(1), \dots, x(N)\}$ of N data objects and a partition of these data into k mutually disjoint cluster, N_i is the number of objects assigned to the i_{th} cluster, $x_l(I)$ is the i_{th} object assigned to that cluster, \bar{x}_i is the n -dimensional vector of sample means within that cluster (cluster centroid) and \bar{x} is the n -dimensional vector of overall sample means (data centroid). As such, the within-group and between-group matrices sum up to the scatter matrix of the data set [33,34]. As a consequence, compact

and separated clusters are expected to have small values of W and large values of B . Hence, the better the data partition the greater the value of the ratio between B and W [34].

Features selection by PCA descriptor

We have defined an iterative method for the selection of the relevant features (among our 40 characteristics) that distinguish better two groups of images. The method selects and evaluates features using the descriptor explained above. The method tests every possible combination of two features and applies the PCA. The method keeps the ten combinations of two features with higher PCA descriptor value. In the second iteration, all features are individually tested again in combination with the ten couples of two features. Again, all the combinations are evaluated and the program keeps the five with higher PCA descriptor value for each one of the ten couples. Therefore, at this moment the program handles 50 trios of features. In the next iteration, the same process is repeated but only two best features are added, accumulating 100 quartets of features. The process continues adding only one feature per iteration step. Finally, the process is stopped when seven features have been selected or when the value for the PCA descriptor is lower than in the previous step. The selected combination of features is the one with a higher PCA descriptor value.

Supporting Information

Figure S1 Randomization of images combinations and their respective PCA descriptor values. Graph showing the number of combinations with a determined PCA descriptor value. Only four combinations present a PCA descriptor value higher than dWL-dWP combination (yellow square). (TIF)

Figure S2 Randomization of images combinations and the MANOVA test p -value for their respective polygon distribution. Graph showing the number of combinations with a determined p -value for the MANOVA test. The polygon distribution for each combination was compared using the MANOVA test. The number of cases with a determined p -value is represented. The yellow square corresponds to the p -value of the combination dWL-dWP. (TIF)

Table S1 Values for the 40 characteristics analyzed in the 31 epithelial images. The table shows both real and normalized quantities after the extraction of the values for the 40 characteristics. Yellow boxes highlight the characteristics selected after the PCA descriptor feature selection step. The bottom part graph compares the normalized values of dWL (green) and dWP (red) for each characteristic. (XLSX)

Text S1 Supporting Definitions. (DOC)

Acknowledgments

We would like to thank L. daF. Costa, F.A. Martin, C. Serrano, B. Acha and members of Dr. Pascual and Dr. Escudero lab for providing helpful comments on this work.

Author Contributions

Conceived and designed the experiments: LME AP. Performed the experiments: DS-G LME. Analyzed the data: LME AS AP DS-G. Wrote the paper: LME AS AP D-SG.

References

- Lecuit T, Le Goff L (2007) Orchestrating size and shape during morphogenesis. *Nature* 450: 189–192.
- Axelrod JD (2006) Cell shape in proliferating epithelia: a multifaceted problem. *Cell* 126: 643–645.
- Escudero LM, Bischoff M, Freeman M (2007) Myosin II regulates complex cellular arrangement and epithelial architecture in *Drosophila*. *Dev Cell* 13: 717–729.
- Bertet C, Sulak L, Lecuit T (2004) Myosin-dependent junction remodelling controls planar cell intercalation and axis elongation. *Nature* 429: 667–671.
- Simoes Sde M, Blankenship JT, Weitz O, Farrell DL, Tamada M, et al. (2010) Rho-kinase directs Bazooka/Par-3 planar polarity during *Drosophila* axis elongation. *Dev Cell* 19: 377–388.
- Leptin M, Grunewald B (1990) Cell shape changes during gastrulation in *Drosophila*. *Development* 110: 73–84.
- Costa M, Wilson ET, Wieschaus E (1994) A putative cell signal encoded by the folded gastrulation gene coordinates cell shape changes during *Drosophila* gastrulation. *Cell* 76: 1075–1089.
- Parks S, Wieschaus E (1991) The *Drosophila* gastrulation gene *concertina* encodes a G alpha-like protein. *Cell* 64: 447–458.
- Kolsch V, Scher T, Fernandez-Ballester GJ, Serrano L, Leptin M (2007) Control of *Drosophila* gastrulation by apical localization of adherens junctions and RhoGEF2. *Science* 315: 384–386.
- Barrett K, Leptin M, Settleman J (1997) The Rho GTPase and a putative RhoGEF mediate a signaling pathway for the cell shape changes in *Drosophila* gastrulation. *Cell* 91: 905–915.
- Pilot F, Lecuit T (2005) Compartmentalized morphogenesis in epithelia: from cell to tissue shape. *Dev Dyn* 232: 685–694.
- Haigo SL, Hildebrand JD, Harland RM, Wallingford JB (2003) Shroom induces apical constriction and is required for hinge point formation during neural tube closure. *Curr Biol* 13: 2125–2137.
- Hildebrand JD, Soriano P (1999) Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice. *Cell* 99: 485–497.
- Milan M, Campuzano S, Garcia-Bellido A (1996) Cell cycling and patterned cell proliferation in the *Drosophila* wing during metamorphosis. *Proc Natl Acad Sci U S A* 93: 11687–11692.
- Martin FA, Herrera SC, Morata G (2009) Cell competition, growth and size control in the *Drosophila* wing imaginal disc. *Development* 136: 3747–3756.
- Farhadifar R, Roper JC, Aigouy B, Eaton S, Julicher F (2007) The influence of cell mechanics, cell-cell interactions, and proliferation on epithelial packing. *Curr Biol* 17: 2095–2104.
- Gibson MC, Patel AB, Nagpal R, Perrimon N (2006) The emergence of geometric order in proliferating metazoan epithelia. *Nature* 442: 1038–1041.
- Gho M, Bellaiche Y, Schweisguth F (1999) Revisiting the *Drosophila* microchaete lineage: a novel intrinsically asymmetric cell division generates a glial cell. *Development* 126: 3573–3584.
- Classen AK, Anderson KI, Marois E, Eaton S (2005) Hexagonal packing of *Drosophila* wing epithelial cells by the planar cell polarity pathway. *Dev Cell* 9: 805–817.
- Martinez-Arias M, Stewart A (2002) *Molecular Principles of Animal Development*: OUP Oxford. 424 p.
- Patel AB, Gibson WT, Gibson MC, Nagpal R (2009) Modeling and inferring cleavage patterns in proliferating epithelia. *PLoS Comput Biol* 5: e1000412.
- Aegerter-Wilmsen T, Smith AC, Christen AJ, Aegerter CM, Hafen E, et al. (2010) Exploring the effects of mechanical feedback on epithelial topology. *Development* 137: 499–506.
- Gibson WT, Gibson MC (2009) Cell topology, geometry, and morphogenesis in proliferating epithelia. *Curr Top Dev Biol* 89: 87–114.
- Eaton S, Julicher F (2011) Cell flow and tissue polarity patterns. *Curr Opin Genet Dev* 21: 747–752.
- Gibson WT, Veldhuis JH, Rubinstein B, Cartwright HN, Perrimon N, et al. (2011) Control of the mitotic cleavage plane by local epithelial topology. *Cell* 144: 427–438.
- Escudero LM, Costa Lda F, Kicheva A, Briscoe J, Freeman M, et al. (2011) Epithelial organisation revealed by a network of cellular contacts. *Nat Commun* 2: 526.
- Costa LD, Rodrigues FA, Travieso G, Boas PRV (2007) Characterization of complex networks: a survey of measurements. *Advances in Physics* 56: 167–242.
- Sáez A, Rivas E, Montero-Sánchez A, Paradas C, Acha B, et al. (2013) Quantifiable diagnosis of muscular dystrophies and neurogenic atrophies through network analysis. *BMC Medicine* 11.
- Aldaz S, Escudero LM, Freeman M (2010) Live imaging of *Drosophila* imaginal disc development. *Proc Natl Acad Sci U S A* 107: 14217–14222.
- Freeman LC (1979) Centrality in social networks: Conceptual clarification. *Social Networks* 1: 215–239.
- Newman M (2005) A measure of betweenness centrality based on random walks. *Social Networks* 27: 39–54.
- Aegerter-Wilmsen T, Heimlicher MB, Smith AC, de Reuille PB, Smith RS, et al. (2012) Integrating force-sensing and signaling pathways in a model for the regulation of wing imaginal disc size. *Development* 139: 3221–3231.
- Fukunaga K (1990) *Introduction to Statistical Pattern Recognition*: Academic Press.
- Costa LD, Cesar RMJ (2000, 2009) *Shape Analysis and Classification: Theory and Practice*: CRC Press. 680 p.

